

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
12. Februar 2004 (12.02.2004)

PCT

(10) Internationale Veröffentlichungsnummer  
WO 2004/013333 A2

- (51) Internationale Patentklassifikation<sup>7</sup>: C12N 15/82, 9/12, 15/11, A01H 5/00
- (21) Internationales Aktenzeichen: PCT/EP2003/007877
- (22) Internationales Anmeldedatum:  
18. Juli 2003 (18.07.2003)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:  
102 34 287.3 26. Juli 2002 (26.07.2002) DE
- (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): BASF PLANT SCIENCE GMBH [DE/DE]; ., 67056 Ludwigshafen (DE).
- (72) Erfinder; und
- (75) Erfinder/Anmelder (nur für US): KOCK, Michael [DE/DE]; Am Leutbusch 12, 67105 Schifferstadt (DE). FRANK, Markus [DE/DE]; Rheindammstr. 30, 68163 Mannheim (DE). BADUR, Ralf [DE/DE]; Theodor-Storm-Str. 7B, 67117 Limburgerhof (DE).
- (74) Anwalt: GOLDSCHIED, Bettina; c/o BASF Aktiengesellschaft, 67056 Ludwigshafen (DE).
- (81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Bestimmungsstaaten (regional): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Veröffentlicht:  
— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts
- Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

BEST AVAILABLE COPY

(54) Title: NOVEL SELECTION METHOD

(54) Bezeichnung: NEUE SELEKTIONSVERFAHREN

(57) Abstract: The invention relates to methods for producing transformed plant cells or organisms by transforming a population of plant cells comprising at least one marker protein having a directly or indirectly toxic effect therefor, by means of at least one nucleic acid sequence to be inserted, said sequence being combined with at least one compound preferably a DNA construct which is able to reduce the expression, quantity, activity and/or function of the marker protein. The transformed plant cells have a growth advantage in relation to the non-transformed cells as a result of the action of said compound.

(57) Zusammenfassung: Die vorliegende Erfindung betrifft Verfahren zur Herstellung transformierter pflanzlicher Zellen oder Organismen durch Transformation einer Population pflanzlicher Zellen, die mindestens ein Markerprotein mit einem für diese direkt oder indirekt toxischen Effekt umfasst, mit mindestens einer zu insertierenden Nukleinsäuresequenz in Kombination mit mindestens einer Verbindung - bevorzugt einem DNA-Konstrukt - befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins, wobei die transformierten pflanzlichen Zellen infolge der Wirkung besagter Verbindung gegenüber nicht-transformierten Zellen einen Wachstumsvorteil haben.

WO 2004/013333 A2

## Neue Selektionsverfahren

## Beschreibung

5

- Die vorliegende Erfindung betrifft Verfahren zur Herstellung transformierter pflanzlicher Zellen oder Organismen durch Transformation einer Population pflanzlicher Zellen, die mindestens ein Markerprotein mit einem für diese direkt oder indirekt
- 10 toxischen Effekt umfasst, mit mindestens einer zu insertierenden Nukleinsäuresequenz in Kombination mit mindestens einer Verbindung - bevorzugt einem DNA-Konstrukt - befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins, wobei die transformierten pflanzlichen Zellen
- 15 infolge der Wirkung besagter Verbindung gegenüber nicht-transformierten Zellen einen Wachstumsvorteil haben.

- Die Einführung genetischen Materials in Zielzellen gelingt meist nur in einer sehr begrenzten Anzahl von Zellen einer Population.
- 20 Dies macht die Unterscheidung und Isolierung von erfolgreich transformierten von nicht-transformierten Zellen erforderlich, ein Verfahren das als Selektion bezeichnet wird. Traditionell erfolgt die Selektion mittels einer sogenannten positiven Selektion, wobei die transformierte Zelle in die Lage versetzt
- 25 wird, zu wachsen und zu überleben, wohingegen die untransformierte Zelle im Wachstum gehemmt oder abgetötet wird (McCormick et al. (1986) Plant Cell Reports 5:81-84). Üblicherweise wird eine derartige positive Selektion durch Gene realisiert, die für eine Resistenz gegen ein Biozid kodieren (z.B.
- 30 ein Herbizid wie Phosphinothricin, Glyphosat oder Bromoxynil, einen Metabolismusinhibitor wie 2-Desoxyglucose-6-phosphat (WO 98/45456) oder ein Antibiotikum wie Tetracyclin, Ampicillin, Kanamycin, G 418, Neomycin, Bleomycin oder Hygromycin). Derartige Gene werden auch als positive Selektionsmarker bezeichnet. Der
- 35 positive Selektionsmarker wird gekoppelt (physikalisch oder mittels Cotransformation) mit der in das Zellgenom einzuführenden Nukleinsäuresequenz in die Zelle eingebracht. Anschließend werden die Zellen auf einem Medium unter dem entsprechenden Selektionsdruck (z.B. in Gegenwart eines entsprechenden Antibiotikums oder
- 40 Herbizids) kultiviert, wodurch die transformierten Zellen aufgrund der erworbenen Resistenz gegen besagten Selektionsdruck einen Wachstums-/Überlebensvorteil haben und so selektioniert werden können. Beispielhaft als positive Selektionsmarker seien genannt:

45

## 2

- Phosphinothricinacetyltransferasen (PAT) (auch Bialophos®-Resistenz; bar) acetylieren die freie Aminogruppe des Glutaminsynthaseinhibitors Phosphinothricin (PPT) und erreichen damit eine Detoxifizierung (de Block et al. (1987) EMBO J 6:2513-2518; Vickers JE et al. (1996) Plant Mol Biol Reporter 14:363-368; Thompson CJ et al. (1987) EMBO J 6:2519-2523).
- 5-Enolpyruvylshikimat-3-phosphatsynthasen (EPSPS) verleihen eine Resistenz gegen das unselektive Herbizid Glyphosat® (N-(Phosphonomethyl)glycin; Steinrücken HC et al. (1980) Biochem Biophys Res Commun 94:1207-1212; Levin JG und Sprinson DB (1964) J Biol Chem 239:1142-1150; Cole DJ (1985) Mode of action of glyphosate; A literature analysis, p. 48-74. In: Grossbard E und Atkinson D (eds.) The herbicide glyphosate. Butterworths, Boston.). Glyphosat-tolerante EPSPS Varianten zur Verwendung als Selektionsmarker sind beschrieben (Padgett SR et al. (1996). New weed control opportunities: development of soybeans with a Roundup Ready™ gene. In: Herbicide Resistant Crops (Duke SO, ed.), pp. 53-84. CRC Press, Boca Raton, FL; Saroha MK und Malik VS (1998) J Plant Biochemistry and Biotechnology 7:65-72; Padgett SR et al. (1995) Crop Science 35(5):1451-1461; US 5,510,471; US 5,776,760; US 5,864,425; US 5,633,435; US 5,627,061; US 5,463,175; EP-A 0 218 571).
- Neomycinphosphotransferasen verleihen eine Resistenz gegen Aminoglykosid-Antibiotika wie Neomycin, G418, Hygromycin, Paromomycin oder Kanamycin, indem sie durch eine Phosphorylierungsreaktion deren inhibierende Wirkung reduzieren (Beck et al. (1982) Gene 19:327-336).
- 2-Desoxyglukose-6-phosphatphosphatasen verleihen eine Resistenz gegen 2-Desoxyglukose (EP-A 0 807 836; Randez-Gil et al. (1995) Yeast 11:1233-1240; Sanz et al. (1994) Yeast 10:1195-1202).
- Acetolactatsynthasen verleihen eine Resistenz gegen Imidazolinon/Sulfonylharnstoff-Herbizide (z.B. Imazamox, Imazapyr, Imazaquin, Imazethapyr, Amidosulfuron, Azimsulfuron, Chlorimuronethyl, Chlorsulfuron; Sathasivan K et al. (1990) Nucleic Acids Res 18(8):2188).

Darüber hinaus sind Resistenzgene gegen die Antibiotika Hygromycin (Hygromycinphosphotransferasen), Chloramphenicol (Chloramphenicolacetyltransferase), Tetracyclin, Streptomycin, Zeocin und Ampicillin ( $\beta$ -Lactamase Gen; Datta N, Richmond MH. (1966) Biochem

J 98(1):204-9) beschrieben.

Gene wie die Isopentenyltransferase (*ipt*) aus *Agrobacterium tumefaciens* (strain:PO22) (Genbank Acc.-No.: AB025109) können ebenfalls als Selektionsmarker eingesetzt werden. Das *ipt* Gen ist ein Schlüsselenzym der Cytokin-Biosynthese. Seine Überexpression erleichtert die Regeneration von Pflanzen (z.B. Selektion auf Cytokin-freiem Medium) (Ebinuma H et al. (2000) Proc Natl Acad Sci USA 94:2117-2121; Ebinuma H et al. (2000) Selection of Marker-free transgenic plants using the oncogenes (*ipt*, *rol A*, *B*, *C*) of *Agrobacterium* as selectable markers, In Molecular Biology of Woody Plants. Kluwer Academic Publishers). Nachteilig ist hier zum einen, dass der Selektionsvorteil auf meist subtilen Unterschieden in der Zellproliferation beruht, zum anderen die Pflanze durch die Transformation mit einem Onkogen unerwünschte Eigenschaften (Galltumorbildung) erhält.

Verschiedene weitere positive Selektionsmarker sind in EP-A 0 601 092 beschrieben. Beispielhaft sind zu nennen:  $\beta$ -Glucuronidase (in Verbindung mit z.B. Cytokininglucuronid), Mannose-6-phosphat-Isomerase (in Verbindung mit Mannose), UDP-Galaktose-4-Epimerase (in Verbindung mit z.B. Galactose).

Negative Selektionsmarker werden zur Selektion von Organismen mit erfolgreich deletierten Markersequenzen eingesetzt (Koprek T et al. (1999) Plant J 19(6):719-726). In Gegenwart eines negativen Selektionsmarkers wird die entsprechende Zelle abgetötet oder erfährt einen Wachstumsnachteil. Bei der negativen Selektion wird beispielsweise durch den in die Pflanze eingebrachten negativen Selektionsmarker eine Verbindung, die ansonsten für die Pflanze keine nachteilige Wirkung hat, in eine Verbindung mit nachteiliger (d.h. toxischer) Wirkung umgesetzt. Beispiele für negative Selektionsmarker umfassen: Thymidinkinase (TK) z.B. des Herpes Simplex Virus (Wigler et al. (1977) Cell 11:223), zelluläre Adeninphosphoribosyltransferase (APRT) (Wigler et al. (1979) Proc Natl Acad Sci USA 76:1373), Hypoxanthinphosphoribosyltransferase (HPRT) (Jolly et al. (1983) Proc Natl Acad Sci USA 80:477), Diphtheria Toxin A Fragment (DT-A), die bakterielle Xanthin-Guaninphosphoribosyltransferase (*gpt*; Besnard et al. (1987) Mol. Cell. Biol. 7:4139; Mzoz and Moolten (1993) Human Gene Therapy 4:589-595), das *codA* Genprodukt kodierend für eine Cytosindesaminase (Gleave AP et al. (1999) Plant Mol Biol. 40(2):223-35; Perera RJ et al. (1993) Plant Mol Biol 23(4): 793-799; Stougaard J; (1993) Plant J 3:755-761; EP-A1 595 873), das Cytochrom P450 Gen (Koprek et al. (1999) Plant J 16:719-726), Gene kodierend für eine Haloalkandehalogenase (Naested H (1999) Plant J 18:571-576), das *iaaH* Gen (Sundaresan V et al. (1995) Genes & Development 9:1797-1810) oder das *tms2* Gen (Fedoroff NV & Smith DL (1993)



Plant J 3: 273-289). Die negativen Selektionsmarker werden meist in Kombination mit sogenannten "Prodrugs" oder "Pro-Toxinen" eingesetzt, Verbindungen die durch die Aktivität des Selektionsmarkers in Toxine umgesetzt werden.

5

5-Methylthioribose-(MTR)-kinase ist ein Enzym, dessen enzymatische Aktivität nicht in Säugern, wohl aber in Pflanzen, Bakterien und Protozoen beschrieben ist. Das Enzym kann ein MTR-Analog (5-(Trifluoromethyl)thioribose) als sogenanntes "subversives

10 Substrat" des Methionin-Ausweich-Stoffwechselwegs ("Salvage Pathway") über ein instabiles Intermediat zu der toxischen Verbindung Carbothionyldifluorid umsetzen.

Besagte Selektionssysteme haben verschiedene Nachteile. Der

15 eingebrachte Selektionsmarker (z.B. Antibiotikaresistenz) hat seine Berechtigung allein während der Transformation und Selektion stellt jedoch später ein in der Regel unnötiges und oft auch unerwünschtes Proteinprodukt dar. Dies kann aus Gründen der Verbraucherakzeptanz und/oder der Zulassung als Lebens-  
20 und/oder Futtermittel unvorteilhaft sein. Nachteilig ist in diesem Zusammenhang ferner, dass der zur Selektion verwendete Selektionsmarker in der Regel genetisch mit der in das Genom zu insertierenden Nukleinsäuresequenz gekoppelt ist und nicht durch Segregation im Rahmen der Vermehrung oder Kreuzung entkoppelt  
25 werden kann. In der Regel ist eine Deletion der Markersequenz erforderlich, was zusätzliche Arbeitsschritte erfordert. Darüberhinaus erfordern biotechnologische Arbeiten in zahlreichen Fällen eine Mehrfachtransformation mit verschiedenen Genkonstrukten. Hier ist für jeden Transformationsschritt ein neuer Selektions-  
30 marker erforderlich, wenn nicht der zuvor verwendete zunächst mühsam deletiert werden soll. Dies erfordert jedoch eine breite Palette gut funktionierender Selektionsmarker, die für die meisten pflanzlichen Organismen nicht zur Verfügung stehen.

35 Es stellte sich folglich die Aufgabe, neue Selektionsverfahren zur Selektion transformierter pflanzlicher Zellen und Organismen bereitzustellen, die möglichst die Nachteile der vorhandenen Systeme nicht mehr aufweisen. Diese Aufgabe wird durch die vorliegende Erfindung gelöst.

40

Ein erster Gegenstand der Erfindung betrifft ein Verfahren zur Herstellung transformierter pflanzlicher Zellen oder Organismen umfassend nachfolgende Schritte:

45 a) Transformation einer Population pflanzlicher Zellen, wobei die Zellen besagter Population mindestens ein Markerprotein enthalten, das für besagte Population direkt oder indirekt

einen toxischen Effekt bewirken kann, mit mindestens einer zu insertierenden Nukleinsäuresequenz in Kombination mit mindestens einer Verbindung befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion mindestens eines Markerproteins, und

5  
10 b) Selektion von transformierten pflanzlichen Zellen, die in ihrem Genom besagte Nukleinsäuresequenz aufweisen und die infolge der Wirkung besagter Verbindung gegenüber nicht-transformierten Zellen einen Wachstumsvorteil haben, aus besagter Population pflanzlicher Zellen, wobei die Selektion unter Bedingungen durchgeführt wird, bei denen das Markerprotein seinen toxischen Effekt auf die nicht transformierten Zellen ausüben kann.

15 In einer bevorzugten Ausführungsform ist das Markerprotein ein Protein, dass in der Lage ist, eine für besagte Population pflanzlicher Zellen nicht-toxische Substanz X in eine für besagte Population toxische Substanz Y direkt oder indirekt umzusetzen.  
20 Das erfindungsgemäße Verfahren umfasst in diesem Fall bevorzugt nachfolgende Schritte:

a) Transformation der Population pflanzlicher Zellen mit mindestens einer zu insertierenden Nukleinsäuresequenz in Kombination mit mindestens einer Verbindung befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion mindestens eines Markerproteins, und  
25  
b) Behandlung besagter Population pflanzlicher Zellen mit der Substanz X in einer Konzentration, die infolge der Umsetzung durch das Markerprotein einen für nicht-transformierte Zellen toxischen Effekt bedingt, und  
30  
c) Selektion von transformierten pflanzlichen Zellen, die in ihrem Genom besagte insertierte Nukleinsäuresequenz aufweisen und die infolge der Wirkung besagter Verbindung gegenüber nicht-transformierten Zellen einen Wachstumsvorteil aufweisen, aus besagter Population pflanzlicher Zellen, wobei die Selektion unter Bedingungen durchgeführt wird, bei denen das Markerprotein seinen toxischen Effekt auf die nicht transformierten Zellen ausüben kann..  
35  
40

Bevorzugt handelt es sich bei der nicht-toxischen Substanz X um eine Substanz, die natürlicherweise in pflanzlichen Zellen oder  
45 Organismen nicht oder nur in Konzentration vorkommt, die im wesentlichen keinen toxischen Effekt bewirken können. Bevorzugt wird die nicht-toxische Substanz X im Rahmen des erfindungs-

gemäßen Verfahrens exogen z.B. über das Medium oder das Wachstumssubstrat appliziert.

- Der Begriff "Verbindung befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion mindestens eines Markerproteins" ist breit zu verstehen und meint allgemein alle Verbindungen, die direkt oder indirekt, alleine oder in Kooperation mit anderen Faktoren eine Verminderung der Proteinmenge, RNA-Menge, Genaktivität, Proteinaktivität - oder funktion mindestens eines Markerproteins bewirken. Besagte Verbindungen sind infolge auch unter der Bezeichnung "anti-Markerprotein"-Verbindungen zusammengefasst. Der Begriff "anti-Markerprotein"-Verbindung schließt insbesondere - jedoch nicht einschränkend - die im Rahmen des erfindungsgemäßen Verfahren in den bevorzugten Ausführungsformen zum Einsatz kommenden Nukleinsäuresequenzen, Ribonukleinsäuresequenzen, doppelsträngigen Ribonukleinsäuresequenzen, antisense-Ribonukleinsäuresequenzen, Expressionskassetten, Peptide, Proteine oder andere Faktoren ein.
- 20 In einer bevorzugten Ausführungsform meint "anti-Markerprotein"-Verbindung ein DNA-Konstrukt umfassend
- a) mindestens eine Expressionskassette geeignet zur Expression einer Ribonukleinsäuresequenz und/oder - gegebenenfalls - eines Proteins, wobei besagte Nukleinsäuresequenz und/oder Protein in der Lage ist, die Expression, Menge, Aktivität und/oder Funktion des Markerproteins zu vermindern, oder
  - b) mindestens eine Sequenz, die eine teilweise oder vollständige Deletion oder Inversion der Sequenz kodierend für besagtes Markerprotein bewirkt und so eine Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins ermöglicht, sowie - gegebenenfalls - weitere Funktionselemente, die besagte Deletion oder Inversion erleichtern und/oder fördern, oder
  - c) mindestens eine Sequenz, die eine Insertion in die Sequenz kodierend für das besagte Markerprotein bewirkt und so eine Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins ermöglicht, sowie - gegebenenfalls - weitere Funktionselemente, die besagte Insertion erleichtern und/oder fördern.
- Das erfindungsgemäße Verfahren hebt die negativ-selektive Wirkung des Markerproteins auf. Insofern wirkt eine "anti-Markerprotein"-Verbindungen direkt (z.B. über die Inaktivierung mittels Insertion in das Gen kodierend für das Markerprotein)

oder indirekt (z.B. mittels der durch die Expressionskassette exprimierten Ribonukleinsäuresequenz und/oder gegebenenfalls des davon translatierte Proteins) als positiver Selektionsmarker. Das erfindungsgemäße Selektionssystem sei infolge als "reverses Selektionssystem" bezeichnet, da es die negativ-selektive Wirkung des Markerproteins "revertiert".

Das erfindungsgemäße Verfahren bedeutet eine sprunghafte Verbreiterung des Repertoires an positiven Selektionsverfahren zur Selektion transformierter pflanzlicher Zellen.

Vorteilhaft ist ferner, dass in bestimmten, bevorzugten Ausführungsform (z.B. durch Wirkung einer doppelsträngigen oder antisense RNA) der Selektionseffekt ohne Expression eines Fremdproteins realisiert werden kann (s.u.).

Vorteilhaft ist darüber hinaus, dass das zur Selektion indirekt verwendete Markerprotein (z.B. der negative Selektionsmarker) nicht genetisch mit der in das Genom zu insertierenden Nukleinsäuresequenz gekoppelt ist. Im Unterschied zu den ansonsten üblichen Selektionsverfahren kann das Markerprotein - wenn es sich um ein Transgen handelt - durch einfache Segregation im Rahmen nachfolgender Vermehrung oder Kreuzung entfernt werden.

"Pflanzliche Zelle" meint im Rahmen der vorliegenden Erfindung jegliche Art von Zelle, die von einem pflanzlichen Organismus abgeleitet oder in diesem vorhanden ist. Der Begriff umfasst dabei beispielhaft Protoplasten, Kallus- oder Zellkulturen, Mikrosporen, Pollen, Zellen in Form von Geweben wie Blättern, Meristem, Blüten, Embryonen, Wurzeln usw. Insbesondere sind all solche Zellen und Zellpopulationen umfasst, die sich als Zielgewebe für eine Transformation eignen.

"Pflanzlicher Organismus" umfasst dabei jeden Organismus, der zur Photosynthese befähigt ist, sowie die von diesem abgeleitete Zellen, Gewebe, Teile oder Vermehrungsgut (wie Samen oder Früchte). Eingeschlossen sind im Rahmen der Erfindung alle Gattungen und Arten höherer und niedrigerer Pflanzen des Pflanzenreiches. Einjährige, mehrjährige, monocotyledone und dicotyledone Pflanzen sowie Gymnospermen sind bevorzugt.

"Pflanze" im Rahmen der Erfindung meint alle Gattungen und Arten höherer und niedrigerer Pflanzen des Pflanzenreiches. Eingeschlossen unter dem Begriff sind die reifen Pflanzen, Saatgut, Sprossen und Keimlinge, sowie davon abgeleitete Teile, Vermehrungsgut (zum Beispiel Knollen, Samen oder Früchte), Pflanzenorgane, Gewebe, Protoplasten, Kallus und andere Kulturen, zum

## 8

- Beispiel Zellkulturen, sowie alle anderen Arten von Gruppierungen von Pflanzenzellen zu funktionellen oder strukturellen Einheiten. Reife Pflanzen meint Pflanzen zu jedem beliebigen Entwicklungsstadium jenseits des Keimlings. Keimling meint eine junge,
- 5 unreife Pflanze in einem frühen Entwicklungsstadium. "Pflanze" umfasst alle einjährigen und mehrjährige, monokotyledonen und dikotyledonen Pflanzen und schließt beispielhaft jedoch nicht einschränkend solche der Gattungen Cucurbita, Rosa, Vitis, Juglans, Fragaria, Lotus, Medicago, Onobrychis, Trifolium,
- 10 Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersicon, Nicotiana, Solarium, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesis, Pelargonium,
- 15 Panieum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Pisum, Phaseolus, Lolium, Oryza, Zea, Avena, Hordeum, Secale, Triticum, Sorghum, Picea und Populus ein.

- Bevorzugt sind Pflanzen nachfolgender Pflanzenfamilien: Amaranthaceae, Asteraceae, Brassicaceae, Carophyllaceae, Chenopodiaceae,
- 20 Compositae, Cruciferae, Cucurbitaceae, Labiatae, Leguminosae, Papilionoideae, Liliaceae, Linaceae, Malvaceae, Rosaceae, Rubiaceae, Saxifragaceae, Scrophulariaceae, Solanaceae, Sterculiaceae, Tetragoniaceae, Theaceae, Umbelliferae.

## 25

- Bevorzugte monokotyle Pflanzen sind insbesondere ausgewählt aus den monokotylen Kulturpflanzen, wie zum Beispiel der Familie der Gramineae wie Alfalfa, Reis, Mais, Weizen oder andere Getreidearten wie Gerste, Hirse, Roggen, Triticale oder Hafer sowie dem
- 30 Zuckerrohr sowie allen Arten von Gräsern.

Bevorzugte dikotyle Pflanzen sind insbesondere ausgewählt aus den dikotylen Kulturpflanzen, wie zum Beispiel

- 35 - Asteraceae wie Sonnenblume, Tagetes oder Calendula und andere mehr,
- Compositae, besonders die Gattung Lactuca, ganz besonders die Art sativa (Salat) und andere mehr,
- 40 - Cruciferae, besonders die Gattung Brassica, ganz besonders die Arten napus (Raps), campestris (Rübe), oleracea cv Tasti (Kohl), oleracea cv Snowball Y (Blumenkohl) und oleracea cv Emperor (Broccoli) und weitere Kohlarten; und der Gattung
- 45 Arabidopsis, ganz besonders die Art thaliana sowie Kresse oder Canola und andere mehr,

## 9

- Cucurbitaceae wie Melone, Kürbis oder Zucchini und andere mehr,
- Leguminosae besonders die Gattung Glycine, ganz besonders die Art max (Sojabohne) sowie Alfalfa, Erbse, Bohnengewächsen oder

5 Erdnuss und andere mehr

- Rubiaceae, bevorzugt der Unterklasse Lamiidae wie beispielsweise Coffea arabica oder Coffea liberica (Kaffestrauch) und andere mehr,

10

- Solanaceae besonders die Gattung Lycopersicon, ganz besonders die Art esculentum (Tomate), die Gattung Solanum, ganz besonders die Art tuberosum (Kartoffel) und melongena (Aubergine) und die Gattung Capsicum, ganz besonders die Art annum (Pfeffer) sowie Tabak und andere mehr,

15

- Sterculiaceae, bevorzugt der Unterklasse Dilleniidae wie beispielsweise Theobroma cacao (Kakaostrauch) und andere mehr,

- 20 - Theaceae, bevorzugt der Unterklasse Dilleniidae wie beispielsweise Camellia sinensis oder Thea sinensis (Teestrauch) und andere mehr,

- Umbelliferae, besonders die Gattung Daucus (ganz besonders die Art carota (Karrotte)) und Apium (ganz besonders die Art graveolens dulce (Sellerie)) und andere mehr;

25

sowie Lein, Baumwolle, Hanf, Flachs, Gurke, Spinat, Möhre, Zuckerrübe und den verschiedenen Baum-, Nuss- und Weinarten,

30 insbesondere Banane und Kiwi.

Pflanzliche Organismen im Sinne der Erfindung sind weiterhin weitere photosynthetisch aktive befähigte Organismen, wie zum Beispiel Algen, Cyanobakterien sowie Moose. Bevorzugte Algen sind

- 35 Grünalgen, wie beispielsweise Algen der Gattung Haematococcus, Phaedactylum tricornatum, Volvox oder Dunaliella. Insbesondere bevorzugt ist Synechocystis.

Besonders bevorzugt ist die Gruppe der Pflanzen bestehend aus

40 Weizen, Hafer, Hirse, Gerste, Roggen, Mais, Reis, Buchweizen, Sorghum, Triticale, Dinkel, Leinsamen, Zuckerrohr, Raps, Kresse, Arabidopsis, Kohllarten, Soja, Alfalfa, Erbse, Bohnengewächsen, Erdnuss, Kartoffel, Tabak, Tomate, Aubergine, Paprika, Sonnenblume, Tagetes, Salat, Calendula, Melone, Kürbis und Zucchini.

45

Am meisten bevorzugt sind

- 5 a) Pflanzen, die zur Ölproduktion geeignet sind, wie beispielsweise Raps, Sonnenblume, Sesam, Färberdistel (*Carthamus tinctorius*), Ölbaum, Soja, Mais, Erdnuss, Rizinus, Ölpalme, Weizen, Kakaotrauch oder verschiedene Nussarten wie beispielsweise Walnuss, Kokosnuss oder Mandel. Unter diesen wieder besonders bevorzugt sind dikotyledonen Pflanzen, insbesondere Raps, Soja und Sonnenblume.
- 10 b) Pflanzen, die der Stärkeproduktion dienen, wie beispielsweise Mais, Weizen oder Kartoffel.
- 15 c) Pflanzen, die als Nahrungs- und/oder Futtermittel und/oder Nutzpflanze genutzt werden und bei denen eine Resistenz gg. Pathogene vorteilhaft wäre, wie beispielsweise Gerste, Roggen, Reis, Kartoffel, Baumwolle, Flachs, Lein.
- 20 d) Pflanzen, die zur Produktion von Feinchemikalien wie beispielsweise Vitaminen und/oder Carotinoiden dienen können, wie beispielsweise Raps.

"Population pflanzlicher Zellen" meint jegliche Gruppe von pflanzlichen Zellen die im Rahmen der vorliegenden Erfindung einer Transformation unterworfen werden kann und von der nach dem 25 erfindungsgemäßen Verfahren transformierte transgene pflanzliche Zellen erhalten und isoliert werden können. Besagte Population kann dabei beispielsweise auch ein pflanzliches Gewebe, Organ oder eine Zellkultur usw. sein. Beispielhaft jedoch nicht einschränkend, kann besagte Population eine isolierte Zygote, ein 30 isolierter unreifer Embryo, embryogener Kallus, Pflänzlichen oder auch verschiedene Blütengewebe (sowohl in vitro als auch in vivo) umfassen.

35 "Genom" meint die Gesamtheit der Erbinformation einer pflanzlichen Zelle und umfasst sowohl die genetische Information des Zellkerns als auch die der Plastiden (z.B. Chloroplasten) und Mitochondrien. Bevorzugt meint Genom jedoch die genetische Information des Zellkerns (beispielsweise der nukleären Chromo- 40 somen).

"Selektion" meint das Identifizieren und/oder Isolieren von erfolgreich transformierten pflanzlichen Zellen aus einer Population nicht-transformierter Zellen unter Einsatz des erfindungs- 45 gemäßen Verfahrens. Dabei ist es nicht zwingend erforderlich, dass die Selektion unmittelbar nach der Transformation direkt mit den transformierten Zellen erfolgt. Es ist auch möglich, die

## 11

- Selektion erst zu einem späteren Zeitpunkt, ja sogar bei einer späteren Generation der aus der Transformation resultierenden pflanzlichen Organismen (bzw. von diesen abgeleiteten Zellen, Geweben, Organen oder Vermehrungsgut) vorzunehmen. So können
- 5 beispielsweise Arabidopsispflanzen direkt mit z.B. der Vakuum-infiltrationsmethode transformiert werden (Clough S & Bent A (1998) Plant J 16(6):735-43; Bechtold N et al. (1993) CR Acad Sci Paris 1144(2):204-212) und ergeben infolge transgene Samen, welche anschließend der Selektion ausgesetzt werden können.
- 10 Die Tatsache, dass die zu insertierende Nukleinsäuresequenz "in Kombination mit" der "anti-Markerprotein"-Verbindung (beispielsweise einem DNA-Konstrukt) transformiert wird, ist breit zu verstehen und meint, dass mindestens eine zu insertierende Nukleinsäuresequenz und mindestens eine "anti-Markerprotein"-Verbindung
- 15 miteinander funktionell gekoppelt sind, so dass das Vorliegen der "anti-Markerprotein"-Verbindung in der pflanzlichen Zelle - und des damit verbundenen Selektionsvorteils - das parallele Vorliegen der insertierten Nukleinsäuresequenz als wahrscheinlich
- 20 anzeigt. Die zu insertierende Nukleinsäuresequenz und die "anti-Markerprotein"-Verbindung (z.B. ein DNA-Konstrukt) können dabei bevorzugt, jedoch nicht zwingend, Teil eines einzigen Nukleinsäurekonstruktes (z.B. eines Transformationskonstruktes oder Transformationsvektors) sein, also physikalisch-chemisch durch
- 25 eine kovalente Bindung gekoppelt vorliegen. Sie können jedoch auch getrennt, beispielsweise im Rahmen einer Co-Transformation, gemeinsam eingeführt werden und auch so ihre Funktion im Rahmen des erfindungsgemäßen Verfahrens wahrnehmen. Im Falle, dass die "anti-Markerproteinverbindung" über die Expression einer RNA
- 30 (beispielsweise eine antisense-RNA oder doppelsträngige RNA) wirkt oder eine solche RNA darstellt, kann "in Kombination" auch solche Ausführungsformen umfassen, bei der die besagte RNA und die von der in das Genom insertierten Nukleinsäuresequenz exprimierte RNA einen RNA-Strang ausbilden.
- 35 "Nicht toxische Substanz X" meint allgemein Substanzen, die im Vergleich zu ihrem Umsetzungsprodukt Y - unter ansonsten gleichen Bedingungen - eine verminderte, bevorzugt eine im wesentlichen fehlende biologische Aktivität - bevorzugt Toxizität - aufweisen. Dabei ist die Toxizität der Substanz Y mindestens doppelt so hoch
- 40 wie die der Substanz X, bevorzugt mindestens fünffach so hoch, besonders bevorzugt mindestens zehnfach so hoch, ganz besonders bevorzugt mindestens zwanzigfach so hoch, am meistens bevorzugt mindestens einhundertfach so hoch. "Gleiche Bedingungen" meint dabei, dass alle Bedingungen abgesehen von den unterschiedlichen
- 45 Substanzen X bzw. Y gleich gehalten werden. Es werden demnach gleiche molare Konzentrationen von X bzw. Y, bei gleichem Medium, Temperatur, Organismenart und -dichte etc. eingesetzt. Die



## 12

Umwandlungen der Substanz X in die Substanz Y kann auf verschiedene Weise z.B. durch Hydrolyse, Deaminierung, Verseifung, Dephosphorylierung, Phosphorylierung, Oxidation oder eine andere Art der Aktivierung, Metabolisierung oder Umsetzung realisiert werden. Beispielfhaft - jedoch nicht einschränkend - kann die Substanz X die inaktive Vorstufe oder Derivat eines Pflanzenwachstumsregulators oder Herbizids sein.

"Toxizität" oder "toxischer Effekt" meint einen messbaren, negativen Einfluss auf die Physiologie der Pflanze oder der pflanzlichen Zelle und kann dabei Symptome wie beispielsweise - jedoch nicht einschränkend - ein vermindertes oder gestörtes Wachstum, eine verminderte oder gestörte Photosyntheserate, eine verminderte oder gestörte Zellteilung, eine verminderte oder gestörte Regeneration einer vollständigen Pflanze aus Zellkultur oder Kallus usw. umfassen.

Die mittels des erfindungsgemässen Verfahrens erfolgreich transformierten pflanzlichen Zellen weisen - anders ausgedrückt - gegenüber den nicht-transformierten Zellen der gleichen Ausgangspopulation einen Wachstumsvorteil oder Selektionsvorteil unter Einwirkung der Substanz "X" auf. Wachstums- oder Selektionsvorteil ist dabei breit zu verstehen und meint beispielsweise die Tatsache, dass besagte transformierte pflanzliche Zellen in der Lage sind, Schösslinge auszubilden und/oder zu vollständigen Pflanzen regenerierbar sind, wohingegen die nicht-transformierten Zellen dies nicht oder nur mit deutlicher Verzögerung realisieren können.

Der Begriff des "Markerproteins" ist breit zu verstehen und meint allgemein all solche Proteine, die befähigt sind,

- i) per se einen toxischen Effekt auf die Pflanze oder pflanzliche Zelle auszuüben, oder
- ii) eine nicht toxische Substanz X in eine für die Pflanze oder pflanzliche Zelle toxische Substanz Y direkt oder indirekt umzusetzen.

Dabei kann es sich bei dem Markerprotein um ein pflanzeigenes, endogenes Gen oder aber auch um ein Transgen aus einem anderen Organismus handeln. Bevorzugt hat das Markerprotein selber keine essentielle Funktion für den das Markerprotein umfassenden Organismus. Übt das Markerprotein per se einen toxischen Effekt aus, so wird es bevorzugt nicht konstitutiv sondern beispielsweise unter einem induzierbaren Promotor exprimiert.

## 13

Bevorzugt setzt jedoch das Markerprotein eine nicht toxische Substanz X in eine für die Pflanze oder pflanzliche Zelle toxische Substanz Y direkt oder indirekt um. Insbesondere bevorzugt sind als Markerprotein die sogenannten "negativen Selektionsmarker", wie sie beispielsweise im Rahmen von gezielten Deletionen aus dem Genom eingesetzt werden.

Beispielhaft jedoch nicht einschränkend sind für Markerproteine zu nennen:

- 10 (a) Cytosindeaminasen (CodA oder CDase), wobei bevorzugt Substanzen wie 5-Fluorocytosin (5-FC) als nicht toxische Substanz X eingesetzt werden. Cytosindeaminasen katalysieren die Deaminierung von Cytosin zu Uracil (Kilstrup M et al. (1989) J Bacteriol 171:2124-2127; Anderson L et al. (1989) Arch Microbiol 152:115-118). Bakterien und Pilze, die CDase-Aktivität aufweisen, konvertieren 5-FC zu dem toxischen Metaboliten ("Y") 5-Fluorouracil (5-FU) (Polak A & Scholer HJ (1975) Chemotherapy (Basel) 21:113-130). 5-FC selbst ist von geringer Toxizität (Bennett JE, in Goodman and Gilman: the Pharmacological Basis of Therapeutics. 8th ed., eds. Gilman AG et al. (Pergamon Press, New York) pp. 1165-1181). 5-FU jedoch hat einen stark zytotoxischen Effekt da es infolge zu Fluoro-UTP (FUTP) und Fluoro-dUMP (FdUMP) metabolisiert wird und so die RNA- und DNA-Synthese inhibiert (Calabrisi P & Chabner BA in Goodman and Gilman: the Pharmacological Basis of Therapeutics. 8th ed., eds. Gilman AG et al. (Pergamon Press, New York) pp. 1209-1263); Damon LE et al. (1989) Pharmac Ther 43:155-189).
- 30 Zellen höherer Pflanzen und Säugerzellen haben keine signifikante CDase-Aktivität und können 5-FC nicht deaminieren (Polak A et al. (1976) Chemotherapy 22:137-153; Koechlin BA et al. (1966) Biochemical Pharmacology 15:434-446). Insofern wird im Rahmen des erfindungsgemäßen Verfahrens die CDase als Transgen (z.B. in Form einer transgenen Expressionskassette) in pflanzliche Organismen eingebracht. Entsprechende transgene pflanzliche Zellen oder Organismen werden dann als Masterpflanzen als Ausgangsmaterial eingesetzt. Entsprechende CDase Sequenzen, transgene pflanzliche Organismen und die Durchführung negativer Selektionsverfahren unter Einsatz von z.B. 5-FC als nicht toxische Substanz X sind dem Fachmann bekannt (WO 93/01281; US 5,358,866; Gleave AP et al. (1999) Plant Mol Biol 40(2):223-35; Perera RJ et al. (1993) Plant Mol Biol 23(4):793-799; Stougaard J (1993) Plant J 3:755-761); EP-A1 595 837; Mullen CA et al. (1992) Proc Natl Acad Sci USA 89(1):33-37; Kobayashi T et al. (1995) Jpn J

## 14

Genet 70(3):409-422; Schlaman HRM & Hooykaas PFF (1997) Plant J 11:1377-1385; Xiaohui Wang H et al. (2001) Gene 272(1-2): 249-255; Koprek T et al. (1999) Plant J 19(6):719-726; Gleave AP et al. (1999) Plant Mol Biol 40(2):223-235; Gallego ME  
 5 (1999) Plant Mol Biol 39(1):83-93; Salomon S & Puchta H (1998) EMBO J 17(20):6086-6095; Thykjaer T et al. (1997) Plant Mol Biol 35(4):523-530; Serino G (1997) Plant J 12(3):697-701; Risseuw E (1997) Plant J 11(4):717-728; Blanc V et al. (1996) Biochimie 78(6):511-517; Corneille S et al.  
 10 (2001) Plant J 27:171-178). Cytosindeaminasen und die dafür kodierenden Gene können aus einer Vielzahl von Organismen, bevorzugt Mikroorganismen, wie beispielsweise den Pilzen *Cryptococcus neoformans*, *Candida albicans*, *Torulopsis glabrata*, *Sporothrix schenckii*, *Aspergillus*, *Cladosporium*  
 15 und *Phialophora* (JE Bennett, Chapter 50: Antifungal Agents, in Goodman and Gilman's the Pharmacological Basis of Therapeutics 8th ed., A.G. Gilman, ed., Pergamon Press, New York, 1990) sowie den Bakterien *E.coli* und *Salmonella typhimurium* (Andersen L et al. (1989) Arch Microbiol 152:115-118)  
 20 erhalten werden.

Auf die im Rahmen der genannten Publikationen offenbarten Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.

25

Insbesondere bevorzugt sind Sequenzen entsprechend der GenBank Acc.-No: S56903, sowie die in EP-A1 595 873 beschriebenen modifizierten *codA* Sequenzen, die eine Expression in Eukaryoten ermöglichen. Bevorzugt sind dabei  
 30 Nukleinsäuresequenzen, die für Polypeptide gemäß SEQ ID NO: 2 oder - bevorzugt - 4 kodieren, insbesondere die Sequenzen gemäß SEQ ID NO: 1 oder - bevorzugt - 3.

(b) Cytochrom P-450 Enzyme, insbesondere das bakterielle Cytochrom P-450 SU1 Genprodukt (CYP105A1) aus *Streptomyces griseolus* (Stamm ATCC 11796), wobei bevorzugt Substanzen wie das Pro-Sulfonylharnstoffherbizid R7402 (2-Methylethyl-2-3-dihydro-N-[(4,6-dimethoxypyrimidin-2-yl)aminocarbonyl]-1,2-benzoisothiazol-7-sulfonamid-1,1-dioxid) als nicht toxische  
 40 Substanz X eingesetzt werden. Entsprechende Sequenzen und die Durchführung negativer Selektionsverfahren unter Einsatz von z.B. R7402 als nicht toxische Substanz X sind dem Fachmann bekannt (O'Keefe DP et al. (1994) Plant Physiol 105:473-482; Tissier AF et al. (1999) Plant Cell 11:1841-1852; Koprek T et al. (1999) Plant J 19(6):719-726; O'Keefe DP (1991) Biochemistry 30(2):447-55). Auf die im Rahmen der genannten  
 45

## 15

Publikationen offenbaren Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.

- 5 Insbesondere bevorzugt sind Sequenzen entsprechend der GenBank Acc.-No: M32238. Bevorzugt sind ferner Nukleinsäuresequenzen, die für das Polypeptid gemäß SEQ ID NO: 6 kodieren, insbesondere die Sequenz gemäß SEQ ID NO: 5.
- 10 (c) Indolessigsäurehydrolasen wie beispielsweise das tms2 Genprodukt aus *Agrobacterium tumefaciens*, wobei bevorzugt Substanzen wie Auxinamidverbindungen oder Naphthalacetamid (NAM) als nicht toxische Substanz X eingesetzt werden (wobei NAM zu Naphthalessigsäure, einer phytotoxischen Substanz, umgesetzt wird). Entsprechende Sequenzen und die Durchführung
- 15 negativer Selektionsverfahren unter Einsatz von z.B. NAM als nicht toxische Substanz X sind dem Fachmann bekannt (Fedoroff NV & Smith DL (1993) Plant J 3:273-289; Upadhyaya NM et al. (2000) Plant Mol Biol Rep 18:227-223; Depicker AG et al. (1988) Plant Cell rep 104:1067-1071; Karlin-Neumann GA
- 20 et al. (1991) Plant Cell 3:573-582; Sundaresan V et al. (1995) Gene Develop 9:1797-1810; Cecchini E et al. (1998) Mutat Res 401(1-2):199-206; Zubko E et al. (2000) Nat Biotechnol 18:442-445). Auf die im Rahmen der genannten Publikationen offenbaren Sequenzen, Materialien und Verfahren wird hiermit
- 25 ausdrücklich Bezug genommen.
- Insbesondere bevorzugt sind Sequenzen entsprechend den GenBank Acc.-No: NC\_003308 (Protein\_id="NP\_536128.1), AE009419, AB016260 (Protein\_id="BAA87807.1) und NC002147. Bevorzugt
- 30 sind ferner Nukleinsäuresequenzen, die für Polypeptide gemäß SEQ ID NO: 8 oder 10 kodieren, insbesondere die Sequenzen gemäß SEQ ID NO: 7 oder 9.
- (d) Haloalkandehalogenasen (dhla Genprodukt) z.B. aus *Xanthobacter autotrophicus* GJ10. Die Dehalogenase hydrolisiert Dihaloalkane wie 1,2-Dichloroethan (DCE) zu halogenierten
- 35 Alkoholen und anorganischen Haliden (Naested H et al. (1999) Plant J 18(5)571-576; Janssen DB et al. (1994) Annu Rev Microbiol 48: 163-191; Janssen DB (1989) J Bacteriol
- 40 171(12):6791-9). Auf die im Rahmen der genannten Publikationen offenbaren Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.

## 16

Insbesondere bevorzugt sind Sequenzen entsprechend der GenBank Acc.-No: M26950. Bevorzugt sind ferner Nukleinsäuresequenzen, die für das Polypeptid gemäß SEQ ID NO: 12 kodieren, insbesondere die Sequenz gemäß SEQ ID NO: 11.

5

- (e) Thymidinkinasen (TK), insbesondere virale TKs aus Viren wie Herpes Simplex Virus, SV40, Cytomegalovirus, Varicella-zoster Virus, insbesondere die TK des Typ 1 Herpes Simplex Virus (TK HSV-1), wobei bevorzugt Substanzen wie Acyclovir, Ganciclovir oder 1,2-Deoxy-2-fluoro- $\beta$ -D-arabinofuranosil-5-iodouracil (FIAU) als nicht toxische Substanz X eingesetzt werden. Entsprechende Sequenzen und die Durchführung negativer Selektionsverfahren unter Einsatz von z.B. Acyclovir, Ganciclovir oder FIAU als nicht toxische Substanz X sind dem Fachmann bekannt (Czako M & Marton L (1994) Plant Physiol 104:1067-1071; Wigler M et al. (1977) Cell 11(1):223-232; McKnight SL et al. (1980) Nucl Acids Res 8(24):5949-5964; McKnight SL et al. (1980) Nucl Acids Res 8(24):5931-5948; Preston et al. (1981) J Virol 38(2):593-605; Wagner et al. (1981) Proc Natl Acad Sci USA 78(3):1441-1445; St. Clair et al. (1987) Antimicrob Agents Chemother 31(6):844-849). Auf die im Rahmen der genannten Publikationen offenbarten Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.

25

- Insbesondere bevorzugt sind Sequenzen entsprechend den GenBank Acc.-No: J02224, V00470 und V00467. Bevorzugt sind ferner Nukleinsäuresequenzen, die für Polypeptide gemäß SEQ ID NO: 14 oder 16 kodieren, insbesondere die Sequenzen gemäß SEQ ID NO: 13 oder 15.

30

- (f) Guaninphosphoribosyltransferasen, Hypoxanthinphosphoribosyltransferasen oder Xanthinguaninphosphoribosyltransferasen, wobei bevorzugt Substanzen wie 6-Thioxanthin oder Allopurinol als nicht toxische Substanz X eingesetzt werden. Bevorzugt sind Guaninphosphoribosyltransferasen (gpt) z.B. aus E. Coli (Besnard et al. (1987) Mol Cell Biol 7:4139; Mzoz and Moolten (1993) Human Gene Therapy 4:589-595; Ono et al. (1997) Hum Gene Ther 8(17):2043-55), Hypoxanthinphosphoribosyltransferasen (HPRT; Jolly et al. (1983) Proc Natl Acad Sci USA 80:477; Fonwick "The HGPRT System", pp. 333-373, M. Gottesman (ed.), Molecular Cell Genetics, John Wiley and Sons, New York, 1985), Xanthinguaninphosphoribosyltransferasen z.B. aus Toxoplasma gondii (Knoll LJ et al. (1998) Mol Cell Biol 18(2):807-814; Donald RG et al. (1996) J Biol Chem 271(24):14010-14019). Auf die im Rahmen der genannten

35

40

45

## 17

Publikationen offenbarten Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.

5 Insbesondere bevorzugt sind Sequenzen entsprechend den GenBank Acc.-No: U10247 (*Toxoplasma gondii* HXGPRT), M13422 (*E. coli* gpt) und X00221 (*E. coli* gpt). Bevorzugt sind ferner Nukleinsäuresequenzen, die für Polypeptide gemäß SEQ ID NO: 18, 20 oder 22 kodieren, insbesondere die Sequenzen gemäß SEQ ID NO: 17, 19 oder 21.

10

(g) Purinnukleosidphosphorylasen (PNP; DeoD Genprodukt) z.B. aus *E. coli*, wobei bevorzugt Substanzen wie 6-Methylpurindeoxyribonukleosid als nicht toxische Substanz X eingesetzt werden. Entsprechende Sequenzen und die Durchführung negativer Selektionsverfahren unter Einsatz von z.B. 6-Methylpurindeoxyribonukleosid als nicht toxische Substanz X sind dem Fachmann bekannt (Sorscher EJ et al. (1994) *Gene Therapy* 1:233-238). Auf die im Rahmen der genannten Publikationen offenbarten Sequenzen, Materialien und Verfahren wird hiermit

15

20 ausdrücklich Bezug genommen.

Insbesondere bevorzugt sind Sequenzen entsprechend der GenBank Acc.-No: M60917. Bevorzugt sind ferner Nukleinsäuresequenzen, die für das Polypeptid gemäß SEQ ID NO: 24 kodieren, insbesondere die Sequenz gemäß SEQ ID NO: 23.

25

h) Phosphonatmonoesterhydrolasen, welche inaktive Esterderivate des Herbizides Glyphosat (z.B. Glycerylglyphosat) zu der aktiven Form des Herbizids umsetzen. Entsprechende Sequenzen und die Durchführung negativer Selektionsverfahren unter Einsatz von z.B. Glycerylglyphosat sind dem Fachmann bekannt (US 5,254,801; Dotson SB et al. (1996) *Plant J* 10(2):383-392; Dotson SB et al. (1996) *J Biol Chem* 271(42): 25754-25761). Auf die im Rahmen der genannten Publikationen offenbarten Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.

30

35

Insbesondere bevorzugt sind Sequenzen entsprechend der GenBank Acc.-No: U44852. Bevorzugt sind ferner Nukleinsäuresequenzen, die für das Polypeptid gemäß SEQ ID NO: 26 kodieren, insbesondere die Sequenz gemäß SEQ ID NO: 25.

40

(i) Aux-1 und - bevorzugt - Aux-2 Genprodukte z.B. der Ti-Plasmide von *Agrobacterium* Stämmen wie *A. rhizogenes* oder *A. tumefaciens* (Beclin C et al. (1993) *Transgenics Res*

45

2:4855); Gaudin V, Jouanin L. (1995) Plant Mol Biol.  
28(1):123-36.

Die Aktivität der beiden Enzyme bedingt die Produktion von  
5 Indolacetamid (IAA) in der pflanzlichen Zelle. Aux-1 kodiert  
für eine Indolacetamidsynthase (IAMS) und setzt Tryptophan zu  
Indolacetamid um (VanOnckelen et al. (1986) FEBS Lett. 198:  
357-360). Aux-2 kodiert für das Enzym Indolacetamidhydrolase  
10 (IAMH) und setzt Indolacetamid, eine Substanz ohne Phyto-  
hormonaktivität, zu dem aktiven Auxin Indolelessigsäure um  
(Inze D et al. (1984) Mol Gen Genet 194:265-274; Tomashow  
et al. (1984) Proc Natl Acad Sci USA 81:5071-5075; Schroder  
et al. (1984) Eur J Biochem 138:387-391). Das Enzym IAMH kann  
15 ferner einer Reihe von Indolamid-Substraten wie beispiels-  
weise Naphthalacetamid hydrolisieren, wobei letzteres in den  
Pflanzenwachstumsregulator Naphthalessigsäure (NAA) umgesetzt  
wird. Die Verwendung des IAMH Gens als negativer Selektions-  
marker ist beispielsweise in US 5,180,873 beschrieben.  
Entsprechende Enzyme sind auch in *A. rhizogenes*, *A. vitis*  
20 (Canaday J et al. (1992) Mol Gen Genet 235:292-303) and  
*Pseudomonas savastanoi* (Yamada et al. (1985) Proc Natl  
Acad Sci USA 82:6522-6526) beschrieben. Der Einsatz als  
negativer Selektionsmarker zum Abtöten bestimmter Zell-  
gewebe (z.B. Pollen; US 5,426,041) oder transgener Pflanzen  
25 (US 5,180,873) ist beschrieben. Entsprechende Sequenzen und  
die Durchführung negativer Selektionsverfahren unter Einsatz  
von z.B. Naphthalacetamid sind dem Fachmann bekannt (s.o.).  
Auf die im Rahmen der genannten Publikationen offenbarten  
Sequenzen, Materialien und Verfahren wird hiermit aus-  
30 drücklich Bezug genommen.

Insbesondere bevorzugt sind Sequenzen entsprechend den Gen-  
Bank Acc.-No: M61151, AF039169 und AB025110. Bevorzugt sind  
35 ferner Nukleinsäuresequenzen, die für Polypeptide gemäß  
SEQ ID NO: 28, 30, 32, 34 oder 36 kodieren, insbesondere  
die Sequenzen gemäß SEQ ID NO: 27, 29, 31, 33 oder 35.

(j) Adeninphosphoribosyltransferasen (APRT), wobei bevorzugt  
40 Substanzen wie 4-Aminopyrazolopyrimidin als nicht toxische  
Substanz X eingesetzt werden. Entsprechende Sequenzen und die  
Durchführung negativer Selektionsverfahren unter Einsatz sind  
dem Fachmann bekannt (Wigler M et al. (1979) Proc Natl Acad  
Sci USA 76(3):1373-6; Taylor et al. "The APRT System", pp.,  
311-332, M. Gottesman (ed.), Molecular Cell Genetics, John  
45 Wiley and Sons, New York, 1985).

## 19

- k) Methoxinindehydrogenasen, wobei bevorzugt Substanzen wie 2-Amino-4-methoxy-butansäure (Methoxinin) als nicht toxische Substanz X eingesetzt werden, welches zum toxischen Methoxy-vinylglycin umgesetzt wird (Margraff R et al. (1980) Experimentia 36: 846).
- l) Rhizobitoxinsynthasen, wobei bevorzugt Substanzen wie 2-Amino-4-methoxy-butansäure (Methoxinin) als nicht toxische Substanz X eingesetzt werden, welches zum toxischen 2-Amino-4-[2-amino-3-hydroxypropyl]-trans-3-butansäure (Rhizobitoxin) umgesetzt wird (Owens LD et al. (1973) Weed Science 21:63-66),
- m) 5-Methylthioribose (MTR) kinasen, wobei bevorzugt Substanzen wie 5-(Trifluoromethyl)thioribose (MTR-Analog, "subversives Substrat") als nicht toxische Substanz X eingesetzt wird, welches über ein instabiles Intermediat zu der toxischen Substanz (Y) Carbothionyldifluorid umgesetzt wird. Die MTR-Kinase ist ein Schlüsselenzym des Methionin-Ausweich-Stoffwechselwegs ("Salvage Pathway"). Entsprechende Enzymaktivitäten wurden nicht in Säugern, wohl aber in Pflanzen, Bakterien und Protozoen beschrieben. MTR Kinasen aus verschiedenen Arten wurden aufgrund definierter Sequenzmotive identifiziert (Sekowska A et al. (2001) BMC Microbiol 1:15; <http://www.biomedcentral.com/1471-2180/1/15>). Entsprechende Sequenzen und die Durchführung negativer Selektionsverfahren unter Einsatz von z.B. 5-(Trifluoromethyl)thioribose sind dem Fachmann bekannt und leicht aus entsprechenden Sequenzdatenbank (z.B. GenBank) erhältlich (Sekowska A et al. (2001) BMC Microbiol 1:15; Cornell KA et al. (1996) 317:285-290). Auf die im Rahmen der genannten Publikationen offenbarten Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.
- Eine pflanzliche MTR-Kinase ist bislang jedoch nicht eindeutig identifiziert worden und wird im Rahmen des erfindungsgemäßen Verfahrens bereitgestellt (SEQ ID NO: 39 bzw. 40). Darüberhinaus werden Homologe aus anderen Pflanzenarten bereitgestellt nämlich aus Mais (SEQ ID NO: 59 bzw. 60), Raps (SEQ ID NO: 61, 63 bzw. 62, 64), Reis (SEQ ID NO: 65 bzw. 66) sowie Soja (SEQ ID NO: 67 bzw. 68).
- Ein weiterer Gegenstand der Erfindung betrifft demnach Aminosäuresequenzen kodieren für eine pflanzliche 5-Methylthioribosekinase, dadurch gekennzeichnet, dass besagte Aminosäuresequenz mindestens eine Sequenz enthält ausgewählt aus der Gruppe bestehend aus SEQ ID NO: 60, 62, 64, 66 oder 68.



Ein weiterer Gegenstand der Erfindung betrifft demnach Nukleinsäuresequenzen kodieren für eine pflanzliche 5-Methylthioribosekinase, dadurch gekennzeichnet, dass besagte

5 Nukleinsäuresequenz mindestens eine Sequenz enthält ausgewählt aus der Gruppe bestehend aus SEQ ID NO: 59, 61, 63, 65 oder 67. Auch wenn besagte Sequenzen teilweise nur Fragmente vollständiger cDNAs darstellen, ist ihre Länge dennoch mehr als ausreichend um eine Verwendung und Funktionalität als

10 antisense-RNA bzw. doppelsträngige RNA zu gewährleisten. Bevorzugt wird als Markerprotein eine pflanzliche, endogene MTR-Kinase verwendet. Weitere endogene pflanzliche MTR-Kinasen können leicht mittels Durchmusterung von Daten- oder Genbanken unter Verwendung konservierter, MTK-Kinase

15 typischer Motive identifiziert werden. Besagte Motive können beispielsweise aus Fig. 9a-b abgeleitet werden. Solche Motive können beispielhaft jedoch nicht einschränkend folgende Sequenzen umfassen:

20 E(V/I)GDGN(L/I)N(L/Y/F)V(F/Y) bevorzugt EVGDGNLN(Y/F)V(F/Y)  
 KQALPY(V/I)RC  
 SWPMT(R/K)ERAYF  
 PEVYHFDRT  
 GMRY(I/L)EPPHI

25 CRLTEQVVFSDPY  
 HGDLH(S/T)GS

Weitere geeignete Motive können unschwer aus Fig.9a-b abgeleitet werden.

30 Insbesondere bevorzugt sind Sequenzen entsprechend den GenBank Acc.-No: AF212863 oder AC079674 (Protein\_ID=AAG51775.1). Bevorzugt sind ferner Nukleinsäuresequenzen, die für Polypeptide gemäß SEQ ID NO: 38 oder 40 kodieren, insbesondere

35 die Sequenzen gemäß SEQ ID NO: 37 oder 39.

n) Alkoholdehydrogenasen (Adh) insbesondere pflanzliche Adh-1 genprodukte, wobei bevorzugt Substanzen wie Allylalkohol als nicht toxische Substanz X eingesetzt werden, welche

40 so zu der toxischen Substanz (Y) Acrolein umgesetzt wird. Entsprechende Sequenzen und die Durchführung negativer Selektionsverfahren unter Einsatz von z.B. Allylalkohol sind dem Fachmann bekannt und leicht aus entsprechenden Sequenzdatenbank (z.B. GenBank) erhältlich (Wisman E et al. (1991)

45 Mol Gen Genet 226(1-2):120-8; Jacobs M et al. (1988) Biochem Genet 26(1-2):105-22; Schwartz D. (1981) Environ Health Perspect 37:75-7). Auf die im Rahmen der genannten Publi-

## 21

kationen offenbaren Sequenzen, Materialien und Verfahren wird hiermit ausdrücklich Bezug genommen.

- Insbesondere bevorzugt sind Sequenzen entsprechend den Gen-Bank Acc.-No: X77943, M12196, AF172282, X04049 oder AF253472.
- 5 Bevorzugt sind ferner Nukleinsäuresequenzen, die für Polypeptide gemäß SEQ ID NO: 42, 44, 46 oder 48 kodieren, insbesondere die Sequenzen gemäß SEQ ID NO: 41, 43, 45 oder 47.
- 10 (o) Weiterhin sind als negativer Selektionsmarker solche Sequenzen geeignet, die per se eine toxische Wirkung auf pflanzliche Zellen ausüben, wie beispielsweise Diphtheriatoxin A, Ribonukleasen wie Barnase sowie Ribosom-inhibierende Proteine wie Ricin. Dabei werden diese Proteine bevorzugt in
- 15 den pflanzlichen Zellen nicht konstitutiv, sondern induzierbar exprimiert. Bevorzugt erfolgt die Induktion chemisch, wobei beispielsweise die unten erwähnten chemisch-induzierbaren Promotoren verwendet werden können, um diese chemisch-induzierte Expression zu gewährleisten.
- 20 "Verminderung" oder "vermindern" ist im Zusammenhang mit einem Markerprotein, bzw. seiner Menge, Expression, Aktivität und/oder Funktion weit auszulegen und umfasst die teilweise oder im wesentlichen vollständige, auf unterschiedliche zell-
- 25 biologische Mechanismen beruhende Unterbindung oder Blockierung der Funktionalität eines Markerproteins in einer pflanzlichen Zelle, Pflanze oder einem davon abgeleiteten Teil, Gewebe, Organ, Zellen oder Samen.
- 30 Eine Verminderung im Sinne der Erfindung umfasst auch eine mengenmäßige Verringerung eines Markerproteins bis hin zu einem im wesentlichen vollständigen Fehlen des Markerproteins (d.h. fehlende Nachweisbarkeit von Markerprotein-Aktivität bzw. Markerprotein-Funktion oder fehlende immunologische Nachweisbarkeit
- 35 des Markerproteins). Dabei wird die Expression eines bestimmten Markerproteins (bzw. seiner Menge, Expression, Aktivität und/oder Funktion) in einer Zelle oder einem Organismus bevorzugt um mehr als 50 %, besonders bevorzugt um mehr als 80 %, ganz besonders bevorzugt um mehr als 90%, am meisten bevorzugt um mehr als
- 40 98 % vermindert. Insbesondere meint Verminderung auch das vollständigen Fehlen des Markerproteins (bzw. seiner Menge, Expression, Aktivität und/oder Funktion). Dabei meint Aktivität und/oder Funktion bevorzugt die Eigenschaft des Markerproteins einen toxischen Effekt auf die pflanzliche Zelle oder den pflanz-
- 45 lichen Organismus auszuüben bzw. die Fähigkeit die Substanz X in die Substanz Y umzusetzen. Bevorzugt wird als der durch das Markerprotein bewirkte toxische Effekt um mehr als 50 %,

besonders bevorzugt um mehr als 80 %, ganz besonders bevorzugt um mehr als 90%, am meisten bevorzugt um mehr als 98 % vermindert. Selbst verständliche umfasst "Verminderung" im Rahmen des erfindungsgemäßen Verfahrens auch eine vollständige, 100%ige Verminderung oder Beseitigung des Markerproteins (bzw. seiner Menge, Expression, Aktivität und/oder Funktion) (z.B. durch Deletion des Markerprotein-Gens aus dem Genom).

Erfindungsgemäß sind verschiedene Strategien zur Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins umfasst. Der Fachmann erkennt, dass eine Reihe verschiedener Methoden zur Verfügung stehen, um die Expression, Menge, Aktivität und/oder Funktion eines Markerproteins in gewünschter Weise zu beeinflussen. Beispielhaft, jedoch nicht einschränkend, seien zu nennen:

- a) Einbringen mindestens einer doppelsträngigen Markerprotein Ribonukleinsäuresequenz (MP-dsRNA) oder einer deren Expression gewährleistenden Expressionskassette oder Expressionskassetten. Umfasst sind solche Verfahren, bei denen die MP-dsRNA gegen ein Markerprotein-Gen (also genomische DNA-Sequenzen wie Promotorsequenzen) oder ein Markerprotein-Gentranskript (also mRNA-Sequenzen) gerichtet ist.
- b) Einbringen mindestens einer Markerprotein antisense-Ribonukleinsäuresequenz (MP-antisenseRNA) oder einer deren Expression gewährleistenden Expressionskassette. Umfasst sind solche Verfahren, bei denen die MP-antisenseRNA gegen ein Markerprotein-Gen (also genomische DNA-Sequenzen) oder ein Markerprotein-Gentranskript (also RNA-Sequenzen) gerichtet ist. Umfasst sind auch  $\alpha$ -anomere Nukleinsäuresequenzen.
- c) Einbringen mindestens einer MP-antisenseRNA kombiniert mit einem Ribozym oder einer deren Expression gewährleistenden Expressionskassette
- d) Einbringen mindestens einer Markerprotein sense-Ribonukleinsäuresequenz (MP-senseRNA) zur Induktion einer Kosuppression oder einer deren Expression gewährleistenden Expressionskassette

## 23

- e) Einbringen mindestens eines DNA-oder Protein-bindenden Faktors gegen ein Markerprotein-Gen, -RNA oder -Protein oder einer dessen Expression gewährleistenden Expressionskassette
- 5 f) Einbringen mindestens einer den Markerprotein RNA-Abbau bewirkenden viralen Nukleinsäuresequenz oder einer deren Expression gewährleistenden Expressionskassette
- 10 g) Einbringen mindestens eines Konstruktes zur Erzeugung eines Funktionsverlustes (z.B. Generierung von Stopp-Kodons, Verschiebungen im Leseraster etc.) an einem Markerprotein-Gen beispielsweise durch Erzeugung einer Insertion, Deletion, Inversion oder Mutation in einem Markerprotein-Gen. Bevorzugt können Knockout-Mutanten mittels gezielter Insertion in
- 15 besagtes Markerprotein-Gen durch homologer Rekombination oder Einbringen von sequenzspezifischen Nukleasen gegen Markerprotein-Gensequenzen generiert werden.
- 20 Dem Fachmann ist bekannt, dass auch weitere Verfahren im Rahmen der vorliegenden Erfindung zur Verminderung eines Markerproteins bzw. seiner Aktivität oder Funktion eingesetzt werden können. Beispielsweise kann auch - je nach Art des verwendeten Markerproteins - das Einbringen einer dominant-negativen Variante eines
- 25 Markerproteins oder einer deren Expression gewährleistenden Expressionskassette vorteilhaft sein. Dabei kann jedes einzelne dieser Verfahren eine Verminderung der Expression, Menge, Aktivität und/oder Funktion eines Markerproteins bewirken. Auch eine kombinierte Anwendung ist denkbar. Weitere Methoden sind dem
- 30 Fachmann bekannt und können die Behinderung oder Unterbindung der Prozessierung des Markerproteins, des Transports des Markerproteins oder dessen mRNA, Hemmung der Ribosomenanlagerung, Hemmung des RNA-Spleißens, Induktion eines Markerprotein-RNA abbauenden Enzyms und/oder Hemmung der Translationselongation
- 35 oder -termination umfassen.

Die einzelnen bevorzugten Verfahren seien infolge durch beispielhafte Ausführungsformen beschrieben:

- 40 a) Einbringen einer doppelsträngigen Ribonukleinsäuresequenz eines Markerproteins (MP-dsRNA)

Das Verfahren der Genregulation mittels doppelsträngiger RNA ("double-stranded RNA interference"; dsRNAi) ist vielfach

- 45 für tierische und pflanzliche Organismen beschrieben (z.B. Matzke MA et al. (2000) Plant Mol Biol 43:401-415; Fire A. et al (1998) Nature 391:806-811; WO 99/32619; WO 99/53050; WO 00/68374;

WO 00/44914; WO 00/44895; WO 00/49035; WO 00/63364). Auf die in den angegebenen Zitaten beschriebenen Verfahren und Methoden wird hiermit ausdrücklich Bezug genommen. dsRNAi-Verfahren beruhen auf dem Phänomen, dass durch gleichzeitiges Einbringen von komplementären Strang- und Gegenstrang eines Gentranskriptes eine hocheffiziente Unterdrückung der Expression des entsprechenden Gens bewirkt wird. Der bewirkte Phänotyp kommt dem einer entsprechenden knock-out Mutanten sehr ähnlich (Waterhouse PM et al. (1998) Proc Natl Acad Sci USA 95:13959-64). Das dsRNAi-Verfahren hat sich bei der Verminderung der Markerprotein-Expression als besonders effizient und vorteilhaft erwiesen.

Doppelsträngiges RNA-Molekül meint im Rahmen der Erfindung bevorzugt eine oder mehr Ribonukleinsäuresequenzen, die aufgrund komplementärer Sequenzen theoretisch (z.B. gemäß den Basenpaarregeln von Waston und Crick) und/oder faktisch (z.B. aufgrund von Hybridisierungsexperimenten in vitro und/oder in vivo) in der Lage sind, doppelsträngige RNA-Strukturen auszubilden. Dem Fachmann ist bewusst, dass die Ausbildung von doppelsträngigen RNA-Strukturen, einen Gleichgewichtszustand darstellt. Bevorzugt ist das Verhältnis von doppelsträngigen Molekülen zu entsprechenden dissoziierten Formen mindestens 1 zu 10, bevorzugt 1:1, besonders bevorzugt 5:1, am meisten bevorzugt 10:1.

Ein weiterer Gegenstand der Erfindung bezieht sich daher auf doppelsträngige RNA-Moleküle (dsRNA-Moleküle), die bei Einbringen in einen pflanzlichen Organismus (oder eine davon abgeleitete Zelle, Gewebe, Organ oder Vermehrungsmaterial) die Verminderung mindestens eines Markerproteins bewirken. Das doppelsträngige RNA-Molekül zur Verminderung der Expression eines Markerproteins (MP-dsRNA) umfasst dabei bevorzugt

- a) einen "sense"-RNA-Strang umfassend mindestens eine Ribonukleotidsequenz, die im wesentlichen identisch ist zu mindestens einem Teil des "sense"-RNA-Transkriptes einer Nukleinsäuresequenz kodierend für ein Markerprotein, und
- b) einen "antisense"-RNA-Strang, der zu dem RNA-"sense"-Strang unter a) im wesentlichen - bevorzugt vollständig - komplementären ist.

In Bezug auf die dsRNA-Moleküle meint Markerprotein-Nukleinsäuresequenz bevorzugt eine Sequenz gemäß SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45 oder 47 oder ein funktionelles Äquivalent derselben.

## 25

- "Im wesentlichen identisch" meint, dass die dsRNA Sequenz auch Insertionen, Deletionen sowie einzelne Punktmutationen im Vergleich zu der Markerprotein Zielsequenz aufweisen kann und dennoch eine effizient Verminderung der Expression bewirkt.
- 5 Bevorzugt beträgt die Homologie (nach weiter unten folgender Definition) mindestens 75 %, bevorzugt mindestens 80 %, ganz besonders bevorzugt mindestens 90 % am meisten bevorzugt 100 % zwischen dem "sense"-Strang einer inhibitorischen dsRNA und mindestens einem Teil des "sense"-RNA-Transkriptes einer Nukleinsäuresequenz kodierend für ein Markerprotein (bzw. zwischen dem
- 10 "antisense"-Strang dem komplementären Strang einer Nukleinsäuresequenz kodierend für ein Markerprotein).

- Eine 100%ige Sequenzidentität zwischen dsRNA und einem Markerprotein Gentranskript ist nicht zwingend erforderlich, um eine effiziente Verminderung der Markerprotein Expression zu bewirken. Demzufolge besteht der Vorteil, dass das Verfahren tolerant ist gegenüber Sequenzabweichungen, wie sie infolge genetischer Mutationen, Polymorphismen oder evolutionärer Divergenzen vor-
- 20 liegen können. So ist es beispielsweise möglich mit der dsRNA, die ausgehend von der Markerprotein Sequenz des einen Organismus generiert wurde, die Markerprotein Expression in einem anderen Organismus zu unterdrücken. Dies ist besonders dann vorteilhaft, wenn als Markerprotein ein pflanzeigenes, endogenes Markerprotein verwendet wird (beispielsweise eine 5-Methylthioribose-
- 25 kinase oder Alkoholdehydrogenase). Zu diesem Zweck umfasst die dsRNA bevorzugt Sequenzbereiche von Markerprotein-Gentranskripten, die konservierten Bereichen entsprechen. Besagte konservierte Bereiche können aus Sequenzvergleichen leicht abgeleitet werden.
- 30

- Die Länge des Teilabschnittes beträgt mindestens 10 Basen, bevorzugt mindestens 25 Basen, besonders bevorzugt mindestens 50 Basen, ganz besonders bevorzugt mindestens 100 Basen, am meisten bevorzugt mindestens 200 Basen oder mindestens 300 Basen.
- 35

- Alternativ, kann eine "im wesentlichen identische" dsRNA auch als Nukleinsäuresequenz definiert werden, die befähigt ist, mit einem Teil eines Markerprotein Gentranskriptes zu hybridisieren (z.B. in 400 mM NaCl, 40 mM PIPES pH 6,4, 1 mM EDTA bei 50°C oder 70°C für 12 bis 16 h).
- 40

- "Im wesentlichen komplementär" meint, dass der "antisense"-RNA-Strang auch Insertionen, Deletionen sowie einzelne Punktmutationen im Vergleich zu dem Komplement des "sense"-RNA-Stranges aufweisen kann. Bevorzugt beträgt die Homologie mindestens 80 %, bevorzugt mindestens 90 %, ganz besonders bevorzugt
- 45

## 26

mindestens 95 %, am meisten bevorzugt 100% zwischen dem "anti-sense"-RNA-Strang und dem Komplement des "sense"-RNA-Stranges.

"Teil des "sense"-RNA-Transkriptes" einer Nukleinsäuresequenz  
5 kodierend für ein Markerprotein meint Fragmente einer RNA oder mRNA transkribiert oder transkribierbar von einer für ein Markerprotein kodierenden Nukleinsäuresequenz, bevorzugt von einem Markerprotein-Gen. Dabei haben die Fragmente bevorzugt eine Sequenzlänge von mindestens 20 Basen, bevorzugt mindestens  
10 50 Basen, besonders bevorzugt mindestens 100 Basen, ganz besonders bevorzugt mindestens 200 Basen, am meisten bevorzugt mindestens 500 Basen. Umfasst ist auch die vollständige transkribierte RNA oder mRNA. Umfasst sind auch Sequenzen wie sie unter künstlichen Bedingungen von Regionen eines Marker-  
15 protein-Gens transkribiert werden können, die ansonsten - unter natürlichen Bedingungen - nicht transkribiert werden, wie beispielsweise Promotorregionen.

Die dsRNA kann aus einem oder mehr Strängen von Polyribonukleo-  
20 tiden bestehen. Natürlich können, um den gleichen Zweck zu erreichen, auch mehrere individuelle dsRNA Moleküle, die jeweils einen der oben definierten Ribonukleotidsequenzabschnitte umfassen, in die Zelle oder den Organismus eingebracht werden. Die doppelsträngige dsRNA-Struktur kann ausgehend von zwei komplementären, separaten RNA-Strängen oder - bevorzugt - ausgehend von  
25 einem einzelnen, selbstkomplementären RNA-Strang gebildet werden. In diesem Fall sind "sense"-RNA-Strang und "antisense"-RNA-Strang bevorzugt kovalent in Form eines invertierten "Repeats" miteinander verbunden.

30 Wie z.B. in WO 99/53050 beschrieben, kann die dsRNA auch eine Haarnadelstruktur umfassen, indem "sense"- und "antisense"-Strang durch eine verbindende Sequenz ("Linker"; beispielsweise ein Intron) verbunden werden. Die selbstkomplementären dsRNA-Struk-  
35 turen sind bevorzugt, da sie lediglich die Expression einer RNA-Sequenz erfordern und die komplementären RNA-Stränge stets in einem äquimolaren Verhältnis umfassen. Bevorzugt kann ist die verbindende Sequenz ein Intron (z.B. ein Intron des ST-LS1 Gens aus Kartoffel; Vancanneyt GF et al. (1990) Mol Gen Genet  
40 220(2):245-250).

Die Nukleinsäuresequenz kodierend für eine dsRNA kann weitere Elemente beinhalten, wie beispielsweise Transkriptionsterminationssignale oder Polyadenylierungssignale.

Sollen die zwei Stränge der dsRNA in einer Zelle oder Pflanze zusammengebracht werden, so kann dies beispielhaft auf folgende Art geschehen:

- 5 a) Transformation der Zelle oder Pflanze mit einem Vektor, der beide Expressionskassetten umfasst,
- b) Kotransformation der Zelle oder Pflanze mit zwei Vektoren, wobei der eine die Expressionskassetten mit dem "sense"-Strang, der andere die Expressionskassetten mit dem "anti-sense"-Strang umfasst.

10

Die Bildung der RNA Duplex kann entweder außerhalb der Zelle oder innerhalb derselben initiiert werden.

15

- Die dsRNA kann entweder in vivo oder in vitro synthetisiert werden. Dazu kann eine DNA-Sequenz kodierend für eine dsRNA in eine Expressionskassette unter Kontrolle mindestens eines genetischen Kontrollelementes (wie beispielsweise einem Promotor) gebracht werden. Eine Polyadenylierung ist nicht erforderlich, ebenso müssen keine Elemente zur Initiierung einer Translation vorhanden sein. Bevorzugt ist die Expressionskassette für die MP-dsRNA auf dem Transformationskonstrukt oder dem Transformationsvektor enthalten. Die Expressionskassetten kodierend für den "antisense"- und/oder den "sense"-Strang einer MP-dsRNA oder für den selbstkomplementären-Strang der dsRNA, werden dazu bevorzugt in einen Transformationsvektor inseriert und mit den unten beschriebenen Verfahren in die pflanzliche Zelle eingebracht. Für das Erfindungsgemäße Verfahren kann eine stabile Insertion in das Genom vorteilhaft sein, ist aber nicht zwingend erforderlich. Da eine dsRNA einen langanhaltenden Effekt bewirkt, ist in vielen Fällen auch eine transiente Expression ausreichend. Die dsRNA kann auch Teil der von der zu insertierenden Nukleinsäuresequenz zu exprimierenden RNA sein, indem sie beispielsweise an den 3'-untranslatierten Teil besagter RNA fusioniert wird.

25

- Die dsRNA kann in einer Menge eingeführt werden, die zumindest eine Kopie pro Zelle ermöglicht. Höhere Mengen (z.B. mindestens 5, 10, 100, 500 oder 1000 Kopien pro Zelle) können ggf. eine effizienter Verminderung bewirken.

30

- b) Einbringen einer antisense-Ribonukleinsäuresequenz eines Markerproteins (MP-antisenseRNA)

- Verfahren zur Verminderung eines bestimmten Proteins durch die "antisense"-Technologie sind vielfach - auch in Pflanzen - beschrieben (Sheehy et al. (1988) Proc Natl Acad Sci USA 85: 8805-8809; US 4,801,340; Mol JN et al. (1990) FEBS Lett

35

40

45



268(2):427-430). Das antisense Nukleinsäuremolekül hybridisiert bzw. bindet mit der zellulären mRNA und/oder genomischen DNA kodierend für das zu vermindernde Markerprotein. Dadurch wird die Transkription und/oder Translation des Markerproteins unter-  
5 drückt. Die Hybridisierung kann auf konventionelle Art über die Bildung einer stabilen Duplex oder - im Fall von genomischer DNA - durch Bindung des antisense Nukleinsäuremoleküls mit der Duplex der genomischen DNA durch spezifische Wechselwirkung in der großen Furche der DNA-Helix entstehen.

- 10 Eine MP-antisenseRNA kann unter Verwendung der für dieses Markerprotein kodierenden Nukleinsäuresequenz, beispielsweise der Nukleinsäuresequenz gemäß SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45  
15 oder 47 nach den Basenpaarregeln von Watson und Crick abgeleitet werden. Die MP-antisenseRNA kann zu der gesamten transkribierten mRNA des Markerproteins komplementär sein, sich auf die kodierende Region beschränken oder nur aus einem Oligonukleotid bestehen, das zu einem Teil der kodierenden oder nicht-kodierenden Sequenz der mRNA komplementär ist. So kann das Oligonukleotid  
20 beispielsweise komplementär zu der Region sein, die den Translationsstart für das Markerprotein umfasst. Die MP-antisenseRNA kann eine Länge von zum Beispiel 5, 10, 15, 20, 25, 30, 35, 40, 45 oder 50 Nukleotide haben, kann aber auch länger sein und mindestens 100, 200, 500, 1000, 2000 oder 5000 Nukleotide umfassen.  
25 MP-antisenseRNA werden im Rahmen des erfindungsgemäßen Verfahrens bevorzugt rekombinant in der Zielzelle exprimiert.

Die MP-antisenseRNA kann auch Teil einer von der zu insertierenden Nukleinsäuresequenz zu exprimierenden RNA sein, indem sie  
30 beispielsweise an den 3'-untranslatierten Teil besagter RNA fusioniert ist.

- Ein weiterer Gegenstand der Erfindung betrifft transgene Expressionskassetten enthaltend eine Nukleinsäuresequenz  
35 kodierend für zumindest einen Teil eines Markerproteins, wobei besagte Nukleinsäuresequenz mit einem in pflanzlichen Organismen funktionellen Promotor in antisense-Orientierung funktionell verknüpft ist. Besagte Expressionskassetten können Teil eines Transformationskonstruktes oder Transformationsvektors sein, oder  
40 aber auch im Rahmen einer Kotransformation eingeführt werden.

In einer weiteren bevorzugten Ausführungsform kann die Expression eines Markerproteins durch Nukleotidsequenzen inhibiert werden, die komplementär zu der regulatorischen Region eines Marker-  
45 protein-Gens (z.B. einem Markerprotein Promoter und/oder Enhancer) sind und triple-helikale Strukturen mit der dortigen DNA-Doppelhelix ausbilden, so dass die Transkription des Markerprotein-Gens vermindert wird. Entsprechende Verfahren sind

beschrieben (Helene C (1991) Anticancer Drug Res 6(6):569-84; Helene C et al. (1992) Ann NY Acad Sci 660:27-36; Maher LJ (1992) Bioassays 14(12):807-815).

- 5 In einer weiteren Ausführungsform kann die MP-antisenseRNA eine  $\alpha$ -anomere Nukleinsäure sein. Derartige  $\alpha$ -anomere Nukleinsäuremoleküle bilden spezifische doppelsträngige Hybride mit komplementärer RNA in denen, - im Unterschied zu den konventionellen  $\beta$ -Nukleinsäuren - die beiden Stränge parallel zueinander verlaufen
- 10 (Gautier C et al. (1987) Nucleic Acids Res 15:6625-6641).

c) Einbringen einer MP-antisenseRNA kombiniert mit einem Ribozym

- Vorteilhaft kann die oben beschriebene antisense-Strategie mit einem Ribozym-Verfahren gekoppelt werden. Katalytische RNA-Moleküle oder Ribozyme können an jede beliebige Ziel-RNA angepasst werden und spalten das Phosphodiester-Gerüst an spezifischen Positionen, wodurch die Ziel-RNA funktionell deaktiviert wird (Tanner NK (1999) FEMS Microbiol Rev 23(3):257-275). Das Ribozym wird dadurch nicht selber modifiziert, sondern ist in der Lage,
- 20 weitere Ziel-RNA-Moleküle analog zu spalten, wodurch es die Eigenschaften eines Enzyms erhält. Der Einbau von Ribozymsequenzen in "antisense"-RNAs verleiht eben diesen "antisense"-RNAs diese enzymähnliche, RNA-spaltende Eigenschaft und steigert so deren Effizienz bei der Inaktivierung der Ziel-RNA.
- 25 Die Herstellung und Verwendung entsprechender Ribozym-"antisense"-RNA-Moleküle ist beschrieben (u.a. bei Haseloff et al. (1988) Nature 334: 585-591); Haselhoff und Gerlach (1988) Nature 334:585-591; Steinecke P et al. (1992) EMBO J 11(4):1525- 1530; de Feyter R et al. (1996) Mol Gen Genet. 250(3):329-338).

- 30 Auf diese Art können Ribozyme (z.B. "Hammerhead"-Ribozyme; Haselhoff und Gerlach (1988) Nature 334:585-591) verwendet werden, um die mRNA eines zu vermindernenden Markerproteins katalytisch zu spalten und so die Translation zu verhindern.
- 35 Die Ribozym-Technologie kann die Effizienz einer antisense-Strategie erhöhen. Verfahren zur Expression von Ribozymen zur Verminderung bestimmter Proteine sind beschrieben in (EP 0 291 533, EP 0 321 201, EP 0 360 257). In pflanzlichen Zellen ist eine Ribozym-Expression ebenfalls beschrieben
- 40 (Steinecke P et al. (1992) EMBO J 11(4):1525-1530; de Feyter R et al. (1996) Mol Gen Genet. 250(3):329-338). Geeignete Zielsequenzen und Ribozyme können zum Beispiel wie bei "Steinecke P, Ribozymes, Methods in Cell Biology 50, Galbraith et al. eds, Academic Press, Inc. (1995), S. 449-460" beschrieben, durch
- 45 Sekundärstrukturberechnungen von Ribozym- und Ziel-RNA sowie durch deren Interaktion bestimmt werden (Bayley CC et al. (1992) Plant Mol Biol. 18(2):353-361; Lloyd AM and Davis RW et al.

## 30

(1994) Mol Gen Genet. 242(6):653-657). Beispielsweise können Derivate der Tetrahymena L-19 IVS RNA konstruiert werden, die komplementäre Bereiche zu der mRNA des zu supprimierenden Markerproteins aufweisen (siehe auch US 4,987,071 und US 5,116,742).

- 5 Alternativ können solche Ribozyme auch über einen Selektionsprozess aus einer Bibliothek diverser Ribozyme identifiziert werden (Bartel D und Szostak JW (1993) Science 261:1411-1418).

- 10 d) Einbringen einer sense-Ribonukleinsäuresequenz eines Markerproteins (MP-senseRNA) zur Induktion einer Kosuppression

Die Expression einer Markerprotein Ribonukleinsäuresequenz (oder eines Teils derselben) in sense-Orientierung kann zu einer Kosuppression des entsprechenden Markerproteingens führen. Die

- 15 Expression von sense-RNA mit Homologie zu einem endogenen Markerproteingen kann die Expression desselben vermindern oder ausschalten, ähnlich wie es für antisense Ansätze beschrieben wurde (Jorgensen et al. (1996) Plant Mol Biol 31(5):957-973; Goring et al. (1991) Proc Natl Acad Sci USA 88:1770-1774; Smith et al. (1990) Mol Gen Genet 224:447-481; Napoli et al. (1990) Plant Cell 2:279-289; Van der Krol et al. (1990) Plant Cell 2:291-99). Dabei kann das eingeführte Konstrukt das zu vermindern, homologe Gen ganz oder nur teilweise repräsentieren. Die Möglichkeit zur Translation ist nicht erforderlich. Die Anwendung dieser
- 20 Technologie auf Pflanzen ist beschrieben (z.B. Napoli et al. (1990) Plant Cell 2:279-289; in US 5,034,323.

- Bevorzugt wird die Kosuppression unter Verwendung einer Sequenz realisiert, die im wesentlichen identisch ist zu zumindest einem
- 30 Teil der Nukleinsäuresequenz kodierend für ein Markerprotein, beispielsweise der Nukleinsäuresequenz gemäß SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45 oder 47.

- 35 Bevorzugt ist die MP-senseRNA so gewählt, dass es nicht zu einer Translation des Markerproteins oder eines Teils desselben kommen kann. Dazu kann beispielsweise der 5'-untranslatierte oder 3'-untranslatierte Bereich gewählt oder aber das ATG-Startkodon deletiert oder mutiert werden.

- 40 e) Einbringen von DNA-oder Protein-bindende Faktoren gegen Markerprotein Gene, -RNAs oder Proteine

- Eine Verminderung einer Markerprotein Expression ist auch mit
- 45 spezifischen DNA-bindenden Faktoren z.B. mit Faktoren vom Typ der Zinkfingertranskriptionsfaktoren möglich. Diese Faktoren lagern sich an die genomische Sequenz des endogenen Zielgens, bevor-

- zugt in den regulatorischen Bereichen, an und bewirken eine Verminderung der Expression. Entsprechende Verfahren zur Herstellung entsprechender Faktoren sind beschrieben (Dreier B et al. (2001) J Biol Chem 276(31):29466-78; Dreier B et al. (2000) J Mol Biol 303(4):489-502; Beerli RR et al. (2000) Proc Natl Acad Sci USA 97 (4):1495-1500; Beerli RR et al. (2000) J Biol Chem 275(42):32617-32627; Segal DJ and Barbas CF 3rd. (2000) Curr Opin Chem Biol 4(1):34-39; Kang JS and Kim JS (2000) J Biol Chem 275(12):8742-8748; Beerli RR et al. (1998) Proc Natl Acad Sci USA 95(25):14628-14633; Kim JS et al. (1997) Proc Natl Acad Sci USA 94(8):3616-3620; Klug A (1999) J Mol Biol 293(2):215-218; Tsai SY et al. (1998) Adv Drug Deliv Rev 30(1-3):23-31; Mapp AK et al. (2000) Proc Natl Acad Sci USA 97(8):3930-3935; Sharrocks AD et al. (1997) Int J Biochem Cell Biol 29(12):1371-1387; Zhang L et al. (2000) J Biol Chem 275(43):33850-33860).

Die Selektion dieser Faktoren kann unter Verwendung eines beliebigen Stückes eines Markerprotein-Gens erfolgen. Bevorzugt liegt dieser Abschnitt im Bereich der Promotorregion. Für eine Genunterdrückung kann er aber auch im Bereich der kodierenden Exons oder Introns liegen.

Ferner können Faktoren in eine Zelle eingebracht werden, die das Markerprotein selber inhibieren. Diese proteinbindenden Faktoren können z.B. Aptamere (Famulok M und Mayer G (1999) Curr Top Microbiol Immunol 243:123-36) oder Antikörper bzw. Antikörperfragmente oder einzelkettige Antikörper sein. Die Gewinnung dieser Faktoren ist beschrieben (Owen M et al. (1992) Biotechnology (N Y) 10(7):790-794; Franken E et al. (1997) Curr Opin Biotechnol 8(4):411-416; Whitelam (1996) Trend Plant Sci 1:286-272).

f) Einbringen von den Markerprotein RNA-Abbau bewirkenden viralen Nukleinsäuresequenzen und Expressionskonstrukten

Die Markerprotein Expression kann effektiv auch durch Induktion des spezifischen Markerprotein RNA-Abbaus durch die Pflanze mit Hilfe eines viralen Expressionssystems (Amplikon; Angell SM et al. (1999) Plant J 20(3):357-362) realisiert werden. Diese Systeme - auch als "VIGS" (viral induced gene silencing) bezeichnet - bringen Nukleinsäuresequenzen mit Homologie zu dem Transkript eines zu vermindernenden Markerproteins mittels viraler Vektoren in die Pflanze ein. Die Transkription wird sodann - vermutlich mediert durch pflanzliche Abwehrmechanismen gegen Viren - abgeschaltet. Entsprechende Techniken und Verfahren sind beschrieben (Ratcliff F et al. (2001) Plant J 25(2):237-45; Fagard M und Vaucheret H (2000) Plant Mol Biol

43(2-3):285-93; Anandalakshmi R et al. (1998) Proc Natl Acad Sci USA 95(22):13079-84; Ruiz MT (1998) Plant Cell 10(6):937-46).

Bevorzugt wird die VIGS-vermittelte Verminderung unter Verwendung  
 5 einer Sequenz realisiert, die im wesentlichen identisch ist zu zumindest einem Teil der Nukleinsäuresequenz kodierend für ein Markerprotein, beispielsweise der Nukleinsäuresequenz gemäß SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45 oder 47.

10

g) Einbringen von Konstrukten zur Erzeugung eines Funktionsverlustes oder einer Funktionsminderung an Markerprotein-Genen

15 Dem Fachmann sind zahlreiche Verfahren bekannt, wie genomische Sequenzen gezielt modifiziert werden können. Dazu zählen insbesondere Verfahren wie die Erzeugung von Knockout-Mutanten mittels gezielter homologen Rekombination z.B. durch Generierung von Stopp-Kodons, Verschiebungen im Leseraster etc. (Hohn B und  
 20 Puchta H (1999) Proc Natl Acad Sci USA 96:8321-8323) oder die gezielte Deletion oder Inversion von Sequenzen mittels z.B. sequenzspezifischer Rekombinasen oder Nukleasen (s.u.)

In einer bevorzugten Ausführungsform wird die Inaktivierung des  
 25 Markerprotein-Gens durch Einbringen einer sequenzspezifischen Rekombinase realisiert. So kann beispielsweise das Markerprotein-Gen Erkennungssequenzen für sequenzspezifische Rekombinasen umfassen oder von solchen flankiert sein, wobei dann durch Einbringen der Rekombinase bestimmte Sequenzen des Markerprotein-  
 30 gens deletiert oder invertiert werden und so eine Inaktivierung des Markerproteinogens erfolgt. Ein entsprechendes Vorgehen ist schematisch in Fig. 1 dargestellt.

Entsprechende Verfahren zur Deletion/Inversion von Sequenzen  
 35 mittels sequenzspezifischer Rekombinasesysteme sind dem Fachmann bekannt. Beispielfhaft seien zu nennen das Cre/lox-System des Bacteriophagen P1 (Dale EC und Ow DW (1991) Proc Natl Acad Sci USA 88:10558-10562; Russell SH et al. (1992) Mol Gen Genet 234:49-59; Osborne BI et al. (1995) Plant J 7:687-701), das FLP/  
 40 FRT System der Hefe (Kilby NJ et al. (1995) Plant J 8:637-652; Lyznik LA et al. (1996) Nucl Acids Res 24:3784-3789), die Gin Rekombinase des Mu Phagen, die Pin Rekombinase aus E.coli oder das R/RS System des pSR1 Plasmids (Onouchi H et al. (1995) Mol Gen Genet 247:653- 660; Sugita Ket al. (2000) Plant J. 22:461-469).  
 45 Bei diesen Systemen interagiert die Rekombinase (beispielsweise Cre oder FLP) spezifisch mit ihren jeweiligen Rekombinationssequenzen (34 bp lox-Sequenz bzw. 47 bp FRT-Sequenz). Bevorzugt

sind das Bacteriophagen Pl Cre/lox und das Hefe FLP/FRT System. Das FLP/FRT und cre/lox Rekombinasesystem wurde bereits in pflanzlichen Systemen angewendet (Odell et al. (1990) Mol Gen Genet 223:369-378). Das Einbringen der Rekombinase wird  
5 bevorzugt mittels rekombinanter Expression ausgehend von einer auf einem DNA-Konstrukt umfassten Expressionskassette realisiert.

Die Verminderung der Markerprotein-Aktivität oder -Menge kann auch durch eine gezielte Deletion im Markerprotein-Gen z.B. durch  
10 sequenzspezifische Induktion von DNA-Doppelstrangbrüchen an einer Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen in oder in der Nähe der Nukleinsäuresequenz kodierend für ein Markerprotein realisiert werden. In seiner einfachsten Ausführungsform (vgl. Fig. 2, A und B) wird hierbei ein Enzym  
15 mit dem Transformationskonstrukt eingebracht, dass mindestens einen Doppelstrangbruch derart erzeugt, dass die resultierende illegitime Rekombination oder Deletion eine Verminderung der Markerprotein-Aktivität oder -Menge - beispielsweise durch Induzieren einer Verschiebung im Leseraster oder Deletion  
20 essentieller Sequenzen - bewirkt.

Die Effizienz dieses Ansatz kann gesteigert werden, indem die Sequenz kodierend für das Markerprotein von Sequenzen (A bzw. A') flankiert ist, die eine ausreichende Länge und Homologie zueinander haben, um infolge des induzierten Doppelstrangbruches miteinander zu rekombinieren und so - durch eine intramolekulare  
25 homologe Rekombination - eine Deletion der Sequenz kodierend für das Markerprotein zu bewirken. Ein entsprechendes Vorgehen ist in einer beispielhaften Ausführungsform dieser Variante in Fig. 3  
30 schematisch dargestellt.

Die Verminderung der Markerprotein-Menge, -Funktion und/oder -Aktivität kann auch durch eine gezielte Insertion von Nukleinsäuresequenzen (z.B. der im Rahmen der erfindungsgemäßen Verfahren  
35 fahrens zu insertierenden Nukleinsäuresequenz) in die Sequenz kodierend für ein Markerprotein (z.B. mittels intermolekularer homologer Rekombination) realisiert werden. Diese Ausführungsform des erfindungsgemäßen Verfahrens ist besonders vorteilhaft und bevorzugt, da es neben den allgemeinen Vorteilen des erfindungs-  
40 gemäßen Verfahrens zudem noch eine reproduzierbare, vorhersagbare, ortsspezifische Insertion der zu insertierenden Nukleinsäuresequenz in das pflanzliche Genom ermöglicht. Dadurch werden die ansonsten im Rahmen einer zufälligen, ortsunspezifischen Insertion auftretenden Positionseffekte (die sich beispielsweise  
45 in Form von unterschiedlichen Expressionshöhen des Transgens oder einer unbeabsichtigten Inaktivierung endogener Gene äußern können) vermieden. Im Rahmen dieser Ausführungsform verwendet man als

- "anti-Markerprotein"-Verbindung bevorzugt ein DNA-Konstrukt, das zumindest einen Teil der Sequenz eines Markerproteingens oder benachbarter Sequenzen umfasst, und so mit diesen in der Zielzelle gezielt rekombinieren kann, so dass durch eine
- 5 Deletion, Addition oder Substitution mindestens eines Nukleotids das Markerproteingens so verändert wird, dass die Funktionalität des Markerprotein-Gens vermindert oder gänzlich aufgehoben wird. Die Veränderung kann auch die regulativen Elemente (z.B. den Promotor) des Markerprotein-Gens betreffen, so dass
- 10 die kodierende Sequenz unverändert bleibt, eine Expression (Transkription und/oder Translation) jedoch unterbleibt und vermindert wird. Bei der konventionellen homologen Rekombination ist die zu insertierende Sequenz an ihrem 5'- und/oder 3'-Ende von weiteren Nukleinsäuresequenzen (A' bzw. B') flankiert, die
- 15 eine ausreichende Länge und Homologie zu entsprechenden Sequenzen des Markerprotein-Gens (A bzw. B) für die Ermöglichung der homologen Rekombination aufweisen. Die Länge liegt in der Regel in einem Bereich von mehreren hundert Basen bis zu mehreren Kilobasen (Thomas KR und Capecchi MR (1987) Cell 51:503; Strepp
- 20 et al. (1998) Proc Natl Acad Sci USA 95(8):4368-4373). Für die homologe Rekombination wird die pflanzliche Zelle mit dem Rekombinationskonstrukt unter Verwendung der unten beschriebenen Verfahren transformiert und erfolgreich rekombinierte Klone basierend auf dem infolge inaktivierten Markerprotein selektio-
- 25 niert. Obgleich homologe Rekombination ein relativ seltenes Ereignis in pflanzlichen Organismen ist, kann durch die Rekombination in das Markerprotein-Gen hinein einem Selektionsdruck ausgewichen werden, was eine Selektion der rekombinierten Zellen und eine hinreichende Effizienz des Verfahrens erlaubt. Ein ent-
- 30 sprechendes Vorgehen ist in einer beispielhaften Ausführungsform dieser Variante in Fig. 4 schematisch dargestellt.

- In einer vorteilhaften Ausführungsform der Erfindung wird jedoch die Insertion in das Markerproteingens mittels weiterer
- 35 Funktionselemente erleichtert. Der Begriff ist umfassend zu verstehen und meint die Verwendung von Sequenzen bzw. von diesen abgeleiteten Transkripten oder Polypeptiden, die die Effizienz der gezielten Integration in ein Markerprotein-Gen zu steigern vermögen. Dazu stehen dem Fachmann verschiedene Verfahren zur
- 40 Verfügung. Bevorzugt wird jedoch die Insertion durch Induktion eines sequenzspezifischen Doppelstrangbruches in oder in der Nähe des Markerprotein-Gens realisiert.

- In einer bevorzugten Ausführungsform der Erfindung wird die
- 45 Inaktivierung (d.h. die Verminderung der Menge, Expression, Aktivität oder Funktion) des Markerproteins durch Integration

einer DNA-Sequenz in ein Markerprotein-Gen realisiert, wobei das Verfahren bevorzugt nachfolgende Schritte umfasst:

- 5 i) Einbringen eines Insertionskonstruktes und mindestens eines Enzyms geeignet zur Induktion von DNA-Doppelstrangbrüchen an einer Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen in oder in der Nähe des Markerprotein-Gens, und
  - 10 ii) Induktion von DNA-Doppelstrangbrüchen an den Erkennungssequenzen zur gezielten Induktion von DNA-Doppelstrangbrüchen in oder in der Nähe des Markerprotein-Gens, und
  - 15 iii) Insertion des Insertionskonstruktes in das Markerprotein-Gen, wobei die Funktionalität des Markerprotein-Gens und - bevorzugt - die Funktionalität der Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen inaktiviert wird, so dass besagte Erkennungssequenz nicht mehr durch das Enzym geeignet zur Induktion von DNA-Doppelstrangbrüchen geschnitten werden kann, und
  - 20 iv) Selektion von Pflanzen oder pflanzlichen Zellen, bei denen das Insertionskonstrukt in das Markerproteingen insertiert wurde.
- 25 Das Insertionskonstrukt umfasst - bevorzugt - die in das Genom zu insertierende Nukleinsäuresequenz, kann aber auch separat von dieser eingesetzt werden.
- "Enzym geeignet zur Induktion von DNA-Doppelstrangbrüchen an der Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen" (infolge "DSBI-Enzym" für "double strand-break inducing enzyme") meint allgemein all solche Enzyme, die in der Lage sind, sequenzspezifisch Doppelstrangbrüche in doppelsträngige DNA zu erzeugen. Beispielsweise, aber nicht einschränkend, sind zu nennen:
- 35 1. Restriktionsendonukleasen, bevorzugt Typ II Restriktionsendonukleasen, besonders bevorzugt Homing-Endonukleasen wie weiter unten im Detail beschrieben.
  - 40 2. Künstliche Nukleasen wie weiter unten im Detail beschrieben, wie beispielsweise chimäre Nukleasen, mutierte Restriktions- oder Homing-Endonukleasen oder RNA-Proteinpartikel abgeleitet von mobilen Introns der Gruppe II.
  - 45 Sowohl natürliche als auch künstlich hergestellte DSBI-Enzyme sind geeignet. Bevorzugt sind all solche DSBI-Enzyme, deren Erkennungssequenz bekannt ist und die entweder in Form ihrer



## 36

Proteine (beispielsweise durch Aufreinigung) gewonnen oder unter Verwendung ihrer Nukleinsäuresequenz exprimiert werden können.

- Bevorzugt wird das DSBII-Enzym unter Kenntnis seiner spezifischen
- 5 Erkennungssequenz so ausgewählt, dass es zusätzlich zu der Ziel-Erkennungssequenz keine weitere funktionellen Erkennungsregionen im Genom der Zielpflanze besitzt. Ganz besonders bevorzugt sind deshalb Homing-Endonukleasen (Übersicht: Belfort M und Roberts RJ (1997) *Nucleic Acids Res* 25:3379-3388; Jasin M (1996) *Trends*
  - 10 *Genet* 12:224-228; Internet: <http://rebase.neb.com/rebase/rebase.homing.html>; Roberts RJ and Macelis D (2001) *Nucl Acids Res* 29: 268-269). Diese erfüllen aufgrund ihrer langen Erkennungssequenzen diese Anforderung. Die für derartige Homing-Endonukleasen kodierenden Sequenzen können beispielsweise aus dem
  - 15 Chloroplastengenom von *Chlamydomonas* isoliert werden (Turmel M et al. (1993) *J Mol Biol* 232:446-467). Geeignete Homing-Endonukleasen sind unter der oben angegebenen Internet-Adresse aufgeführt. Zu nennen sind beispielsweise Homing-Endonukleasen wie F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-CeuI,
  - 20 I-CeuAIIP, I-ChuI, I-CmoI, I-CpaI, I-CpaII, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-CsmI, I-CvuI, I-CvuAIP, I-DdiII, I-DirI, I-DmoI, I-HspNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI,
  - 25 I-PorIIP, I-PpbIP, I-PpoI, I-SPBetaIP, I-ScaI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-SexIP, I-SneIP, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiS3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPA1P, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI,
  - 30 PI-MtuHIP, PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-PspI, PI-Rma43812IP, PI-SPBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII. Bevorzugt sind dabei die Homing-Endonukleasen, deren Gensequenzen bereits bekannt sind, wie beispielsweise F-SceI, I-CeuI, I-ChuI, I-DmoI, I-CpaI, I-CpaII,
  - 35 I-CreI, I-CsmI, F-TevI, F-TevII, I-TevI, I-TevII, I-AniI, I-CvuI, I-LlaI, I-NanI, I-MsoI, I-NitI, I-NjaI, I-PakI, I-PorI, I-PpoI, I-ScaI, I-Ssp6803I, PI-PkoI, PI-PkoII, PI-PspI, PI-TfuI, PI-TliI.

Ganz besonders bevorzugt sind

40

- I-CeuI (Cote MJ und Turmel M (1995) *Curr Genet* 27:177-183.; Gauthier A et al. (1991) *Curr Genet* 19:43-47; Marshall (1991) *Gene* 104:241-245; GenBank Acc.-No.: Z17234 Nukleotide 5102 bis 5758),

45

## 37

- I-ChuI (Cote V et al. (1993) Gene 129:69-76; GenBank Acc.-No.: L06107, Nukleotide 419 bis 1075),
- I-CmoEI (Drouin M et al. (2000) Nucl Acids Res 28:4566-4572),
- I-CpaI aus *Chlamydomonas pallidostigmatica* (GenBank Acc.-No.: L36830, Nukleotide 357 bis 815; Turmel M et al. (1995) Nucleic Acids Res 23:2519-2525; Turmel, M et al. (1995) Mol Biol Evol 12:533-545)
- I-CpaII (Turmel M et al. (1995) Mol Biol Evol 12:533-545; GenBank Acc.-No.: L39865, Nukleotide 719 bis 1423),
- I-CreI (Wang J et al. (1997) Nucleic Acids Res 25: 3767-3776; Dürrenberger, F und Rochaix JD (1991) EMBO J 10:3495-3501; GenBank Acc.-No.: X01977, Nukleotide 571 bis 1062),
- I-CsmI (Ma DP et al. (1992) Plant Mol Biol 18:1001-1004)
- I-NanI (Elde M et al. (1999) Eur J Biochem. 259:281-288; GenBank Acc.-No.: X78280, Nukleotide 418 bis 1155),
- I-NitI (GenBank Acc.-No.: X78277, Nukleotide 426 bis 1163),
- I-NjaI (GenBank Acc.-No.: X78279, Nukleotide 416 bis 1153),
- I-PpoI (Muscarella DE und Vogt VM (1989) Cell 56:443-454; Lin J und Vogt VM (1998) Mol Cell Biol 18:5809-5817; GenBank Acc.-No.: M38131, Nukleotide 86 bis 577),
- I-PspI (GenBank Acc.-No.: U00707, Nukleotide 1839 bis 3449),
- I-ScaI (Monteilhet C et al. (2000) Nucleic Acids Res 28: 1245-1251; GenBank Acc.-No.: X95974, Nukleotide 55 bis 465)
- I-SceI (WO 96/14408; US 5,962,327 dort Seq ID NO: 1),
- Endo SceI (Kawasaki et al. (1991) J Biol Chem 266:5342-5347, identisch zu F-SceI; GenBank Acc.-No.: M63839, Nukleotide 159 bis 1589),
- I-SceII (Sarguiel B et al. (1990) Nucleic Acids Res 18:5659-5665),

- I-SceIII (Sarguiel B et al. (1991) Mol Gen Genet. 255:340-341),
- I-Ssp6803I (GenBank Acc.-No.: D64003, Nukleotide 35372 bis 35824),
- I-TevI (Chu et al. (1990) Proc Natl Acad Sci USA 87:3574-3578; Bell-Pedersen et al. (1990) Nucleic Acids Res 18:3763-3770; GenBank Acc.-No.: AF158101, Nukleotide 144431 bis 143694),
- I-TevII (Bell-Pedersen et al. (1990) Nucleic Acids Res 18:3763-3770; GenBank Acc.-No.: AF158101, Nukleotide 45612 bis 44836),
- I-TevIII (Eddy et al. (1991) Genes Dev. 5:1032-1041),

Ganz besonders bevorzugt sind kommerziell erhältliche Homing-Endonukleasen wie I-CeuI, I-SceI, I-PpoI, PI-PspI oder PI-SceI. Am meisten bevorzugt sind I-SceI und I-PpoI. Während das Gen kodierend für I-PpoI in der natürlichen Form genutzt werden kann, besitzt das Gen kodierend für I-SceI eine Editierstelle. Da die entsprechende Editierung in höherer Pflanzen im Unterschied zu den Mitochondrien der Hefe nicht durchgeführt wird, muss eine künstliche das I-SceI Protein kodierende Sequenz zur heterologen Expression dieses Enzyms eingesetzt werden (US 5,866,361).

Die Enzyme können in der dem Fachmann geläufigen Art und Weise aus ihren Herkunftsorganismen aufgereinigt und/oder die für sie kodierende Nukleinsäuresequenz kloniert werden. Die Sequenzen verschiedener Enzyme sind in der GenBank hinterlegt (s.o.).

Als künstliche DSBI-Enzyme seien beispielhaft chimäre Nukleasen zu nennen, die sich aus einer unspezifischen Nukleasedomäne und einer sequenzspezifischen DNA-Bindungsdomäne (z.B. bestehend aus Zinkfingern) zusammensetzen (Smith J et al. (2000) Nucl Acids Res 28(17):3361-3369; Bibikova M et al. (2001) Mol Cell Biol. 21:289-297). So wurde beispielsweise die katalytische Domäne der Restriktionsendonuklease FokI an Zinkfingerbinde-Domänen fusioniert, wodurch die Spezifität der Endonuklease definiert wurde (Chandrasegaran S & Smith J (1999) Biol Chem 380:841-848; Kim YG & Chandrasegaran S (1994) Proc Natl Acad Sci USA 91:883-887; Kim YG et al. (1996) Proc Natl Acad Sci USA 93:1156-1160). Auch die katalytische Domäne der Ho-Endonuklease aus Hefe konnte bereits mit der beschriebenen Technik eine vordefinierte Spezifität verliehen werden, indem diese an die Zinkfingerdomäne von Transkriptionsfaktoren fusioniert wurde (Nahon E

& Raveh D (1998) Nucl Acids Res 26:1233-1239). Durch geeignete Mutations- und Selektionsverfahren kann man bestehende Homing-Endonukleasen an jede gewünschte Erkennungssequenz anpassen.

- 5 Wie erwähnt, eignen sich insbesondere Zinkfingerproteine als DNA-Bindungsdomäne im Rahmen von chimären Nukleasen. Diese DNA-bindenden Zinkfingerdomänen können an jede beliebige DNA-Sequenz angepasst werden. Entsprechende Verfahren zur Herstellung entsprechender Zinkfingerdomänen sind beschrieben und dem
  - 10 Fachmann bekannt (Beerli RR et al. (2000) Proc Natl Acad Sci 97(4):1495-1500; Beerli RR et al. (2000) J Biol Chem 275(42):32617-32627; Segal DJ and Barbas CF 3rd. (2000) Curr Opin Chem Biol 4(1):34-39; Kang JS and Kim JS (2000) J Biol Chem 275(12):8742-8748; Beerli RR et al. (1998) Proc Natl Acad Sci USA 95(25):14628-14633; Kim JS et al. (1997) Proc Natl Acad Sci USA 94(8):3616-3620; Klug A (1999) J Mol Biol 293(2):215-218; Tsai SY et al. (1998) Adv Drug Deliv Rev 30(1-3):23-31; Mapp AK et al. (2000) Proc Natl Acad Sci USA 97(8):3930-3935; Sharrocks AD et al. (1997) Int J Biochem Cell Biol 29(12):1371-1387; Zhang L et al. (2000) J Biol Chem 275(43):33850-33860). Verfahren zur Herstellung und Selektion von Zink-Finger DNA-Bindedomänen mit hoher Sequenzspezifität sind beschrieben (WO 96/06166, WO 98/53059, WO 98/53057). Durch Fusion einer so erhaltenen DNA-Bindedomäne an die katalytische Domäne einer Endonuklease
  - 25 (wie beispielsweise der FokI oder Ho-Endonuklease) kann man chimäre Nukleasen mit jeder beliebigen Spezifität herstellen, die als DSBI-Enzyme im Rahmen der vorliegenden Erfindung vorteilhaft eingesetzt werden können.
  - 30 Künstliche DSBI-Enzyme mit veränderter Sequenzspezifität können auch durch Mutagenese bereits bekannter Restriktionsendonukleasen oder Homing Endonukleasen durch dem Fachmann geläufige Methoden erzeugt werden. Insbesondere von Interesse ist neben der Mutagenese von Homing-Endonukleasen mit dem Ziel, eine veränderte
  - 35 Substratspezifität zu erzielen, auch die Mutagenese von Maturasen. Maturasen haben häufig viele Gemeinsamkeiten mit Homing-Endonukleasen und lassen sich durch wenige Mutationen ggf. in Nukleasen umwandeln. Dies wurde beispielsweise für die Maturase im bi2 Intron der Bäckerhefe gezeigt. Lediglich zwei Mutationen
  - 40 in dem die Maturase kodierenden offenen Leseraster (ORF) reichten aus, diesem Enzym eine Homing-Endonuklease Aktivität zu verleihen (Szczepanek & Lazowska (1996) EMBO J 15:3758-3767).
- Weitere künstliche Nukleasen können mit Hilfe mobiler Gruppe II
- 45 Introns und den von ihnen kodierten Proteinen oder Teile dieser Proteine erzeugt werden. Mobile Gruppe II Introns bilden mit den von ihnen kodierten Proteinen RNA-Protein-Partikel, die

sequenzspezifisch DNA erkennen und schneiden können. Die Sequenzspezifität kann dabei durch Mutagenese von bestimmten Bereichen des Introns (siehe unten) den Bedürfnissen angepasst werden (WO 97/10362).

5

Bevorzugt wird das DSBI-Enzym als Fusionsprotein mit einer Kernlokalisationssequenz (NLS) exprimiert. Diese NLS-Sequenz ermöglicht einen erleichterten Transport in den Kern und steigert die Effizienz des Rekombinationssystems. Verschiedene NLS-Sequenzen

- 10 sind dem Fachmann bekannt und unter anderem beschrieben bei Jicks GR und Raikhel NV (1995) Annu. Rev. Cell Biol. 11:155-188. Bevorzugt für pflanzliche Organismen ist beispielsweise die NLS-Sequenz des SV40 "large antigen". Ganz besonders bevorzugt sind die nachfolgenden NLS-Sequenzen:

15

NLS1: N-Pro-Lys-Thr-Lys-Arg-Lys-Val-C

NLS2: N-Pro-Lys-Lys-Lys-Arg-Lys-Val-C

- 20 Aufgrund der geringen Größe vieler DSBI-Enzyme (wie beispielsweise der Homing-Endonukleasen) ist jedoch eine NLS-Sequenz nicht zwingend erforderlich. Diese Enzyme können die Kernporen auch ohne die Unterstützung passieren.

- 25 "Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen" meint allgemein solche Sequenzen, die unter den Bedingungen in der jeweils verwendeten eukaryotischen Zelle oder Organismus die Erkennung und Spaltung durch das DSBI-Enzym erlauben. Beispielfhaft aber nicht einschränkend
- 30 seien dabei in nachfolgender Tabelle 1 die Erkennungssequenzen für die jeweiligen aufgeführten DSBI-Enzyme genannt.

Tabelle 1: Erkennungssequenzen und Herkunftsorganismus von DSBI-Enzymen ("^" gibt innerhalb einer Erkennungssequenz die Schnittstelle des DSBI-Enzyms an)

35

DSBI-Enzym	Herkunftsorganismus	Erkennungssequenz
40 CRE	Bacteriophage P1	5'-AACTCTCATCGCTTCGGATAACTTCCTGTTATCCGAAACATATCACTCACTTTGGTGATTTACCGTAAGTGTCTATGATTAATG-3'
FLP	Saccharomyces cerevisiae	5'-GAAGTTCTATTCCGAAGTTCCTATTCTCTAGAAAGTA-TAGGAAGTTC-3'
R	pSR1 Plasmids	5'-CGAGATCATATCACTGTGGACGTTGATGAAAGAATACGTATTCTTTCATCAAATCGT
45 P-Element Transposase	Drosophila	5'-CTAGATGAAATAACATAAGGTGG

DSBI-Enzym	Herkunfts- organismus	Erkennungssequenz
I-AniI	Aspergillus nidulans	5'-TTGAGGAGGTT^TCTCTGTAAATAANNNNNNNNNNNNNNNN 3'-AACTCCTCCAAAGAGACATTTAT^TNNNNNNNNNNNNNNNN^
5 I-DdiI	Dictyostelium discoideumAX3	5'-TTTTTTGGTCATCCAGAAGTATAT 3'-AAAAAACCAG^TAGGTCTTCATATA
I-CvuI	Chlorella vulgaris	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTCAG^CACTCTGTCAAACC
I-CsmI	Chlamydomonas smithii	5'-GTACTAGCATGGGGTCAAATGTCTTTCTGG
10 I-CmoI	Chlamydomo- nasmoewusii	5'-TCGTAGCAGCT^CACGGTT 3'-AGCATCG^TCGAGTGCCAA
I-CreI	Chlamydomonas reinhardtii	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTCAG^CACTCTGTCAAACC
15 I-ChuI	Chlamydomonas humicola	5'-GAAGGTTTGGCACCTCG^ATGTCGGCTCATC 3'-CTTCCAAACCGTG^GAGCTACAGCCGAGTAG
I-CpaI	Chlamydomonas pallidostig- matica	5'-CGATCCTAAGGTAGCGAA^ATTCA 3'-GCTAGGATTCCATC^GCTTTAAGT
I-CpaII	Chlamydomonas pallidostig- matica	5'-CCCGGCTAACTC^TGTGCCAG 3'-GGGCCGAT^TGAGACACGGTC
20 I-CeuI	Chlamydomonas eugametos	5'-CGTAAC^TATAACGGTCCTAA^GGTAGCGAA 3'-GCATTGATATTGCCAG^GATTCCATCGCTT
I-DmoI	Desulfurococ- cus mobilis	5'-ATGCCTTGCCGGGTAA^GTTCCGGCGCGCAT 3'-TACGGAACGGCC^CATTCAGGCCGCGCGTA
25 I-SceI	S.cerevisiae	5'-AGTTACGCTAGGGATAA^CAGGGTAATATAG 3'-TCAATGCGATCCC^TATTGTCCCATATATC 5'-TAGGGATAA^CAGGGTAAT 3'-ATCCC^TATTGTCCCATTA ("Core"-Sequenz)
I-SceII	S.cerevisiae	5'-TTTTGATTCTTTGGTCACCC^TGAAGTATA 3'-AAACTAAGAAACCAG^TGGGACTTCATAT
30 I-SceIII	S.cerevisiae	5'-ATTGGAGGTTTTGGTAAC^TATTTATTACC 3'-TAACCTCCAAAACC^ATTGATAAATAATGG
I-SceIV	S.cerevisiae	5'-TCTTTTCTCTTGATTA^GCCCTAATCTACG 3'-AGAAAAGAGAAC^TAATCGGGATTAGATGC
35 I-SceV	S.cerevisiae	5'-AATAATTTTCT^TCTTAGTAATGCC 3'-TTATTAAAGAAGAATCATTA^CGG
I-SceVI	S.cerevisiae	5'-GTTATTTAATG^TTTTAGTAGTTGG 3'-CAATAAATTACAAAATCATCA^ACC
I-SceVII	S.cerevisiae	5'-TGTCACATTGAGGTGCACTAGTTATTAC
40 PI-SceI	S.cerevisiae	5'-ATCTATGTCGGGTGC^GGAGAAAGAGGTAAT 3'-TAGATACAGCC^CACGCCTCTTTCTCCATTA
F-SceI	S.cerevisiae	5'-GATGCTGTAGGC^ATAGGCTTGGTT 3'-CTACGACA^TCCGTATCCGAACCAA
F-SceII	S.cerevisiae	5'-CTTCCGCAACA^GTAAAATT 3'-GAAAGGCG^TTGTCATTTTAA
45 I-HmuI	Bacillus sub- tilis bacte- riophage SP01	5'-AGTAATGAGCCTAACGCTCAGCAA 3'-TCATTACTCGGATTGC^GAGTCGTT

	DSBI-Enzym	Herkunfts- organismus	Erkennungssequenz
5	I-HmuII	Bacillus subtilis bacteriophage SP82	5'-AGTAATGAGCCTAACGCTCAACAANNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
	I-LlaI	Lactococcus lactis	5'-CACATCCATAAC^CATATCATTTTT 3'-GTGTAGGTATTGGTATAGTAA^AAA
	I-MsoI	Monomastix species	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTGCAG^CACTCTGTCAAACC
10	I-NanI	Naegleria andersoni	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
	I-NitI	Naegleria italica	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
	I-NjaI	Naegleria jamiesoni	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
15	I-PakI	Pseudendoclonium akinetum	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTGCAG^CACTCTGTCAAACC
	I-PorI	Pyrobaculum organotrophum	5'-GCGAGCCCGTAAGGGT^GTGTACGGG 3'-CGCTCGGGCATT^CCCACACATGCCC
20	I-PpoI	Physarum polycephalum	5'-TAACTATGACTCTCTTAA^GGTAGCCAAAT 3'-ATTGATACTGAGAG^AATTCATCGGTTTA
	I-ScaI	Saccharomyces capensis	5'-TGTCACATTGAGGTGCACT^AGTTATTAC 3'-ACAGTGTAACTCCAC^GTGATCAATAATG
	I-Ssp6803I	Synechocystis species	5'-GTCGGGCT^CATAACCCGAA 3'-CAGCCCGAGTA^TTGGGCTT
25	PI-PfuI	Pyrococcus furiosus Vc1	5'-GAAGATGGGAGGAGGG^ACCGGACTCAACTT 3'-CTTCTACCCTCC^TCCCTGGCCTGAGTTGAA
	PI-PfuII	Pyrococcus furiosus Vc1	5'-ACGAATCCATGTGGAGA^AGAGCCTCTATA 3'-TGCTTAGGTACAC^CTCTTCTCGGAGATAT
	PI-PkoI	Pyrococcus kodakaraensis KOD1	5'-GATTTTAGAT^CCCTGTACC 3'-CTAAAA^TCTAGGGACATGG
30	PI-PkoII	Pyrococcus kodakaraensis KOD1	5'-CAGTACTACG^GTTAC 3'-GTCATG^ATGCCAATG
	PI-PspI	Pyrococcus sp.	5'-AAAATCCTGGCAAACAGCTATTAT^GGGTAT 3'-TTTTAGGACCGTTTGTGCGAT^AATACCCATA
35	PI-TfuI	Thermococcus fumicolans ST557	5'-TAGATTTTAGGT^CGCTATATCCTTCC 3'-ATCTAAAA^TCCAGCGATATAGGAAGG
	PI-TfuII	Thermococcus fumicolans ST557	5'-TAYGCNGAYACN^GACGGYTTYT 3'-ATRCGNCT^RTGNCTGCCRAARA
40	PI-ThyI	Thermococcus hydrothermalis	5'-TAYGCNGAYACN^GACGGYTTYT 3'-ATRCGNCT^RTGNCTGCCRAARA
	PI-TliI	Thermococcus litoralis	5'-TAYGCNGAYACNGACGG^YTTYT 3'-ATRCGNCTRTGNC^TGCCRAARA
45	PI-TliII	Thermococcus litoralis	5'-AAATTGCTTGCAAACAGCTATTACGGCTAT
	I-TevI	Bacteriophage T4	5'-AGTGGTATCAAC^GCTCAGTAGATG 3'-TCACCATAGT^TGCGAGTCATCTAC

DSBI-Enzym	Herkunfts- organismus	Erkennungssequenz
I-TevII	Bacteriophage T4	5'-GCTTATGAGTATGAAGTGAACACGT^TATTC 3'-CGAATACTCATACTTCACTTGTG^CAATAAG
5 F-TevI	Bacteriophage T4	5'-GAAACACAAGA^AATGTTTAGTAAANNNNNNNNNNNNNNN 3'-CTTTGTGTTCTTTACAAATCATTTNNNNNNNNNNNNNNNN^
F-TevII	Bacteriophage T4	5'-TTTAATCCTCGCTTC^AGATATGGCAACTG 3'-AAATTAGGAGCGA^AGTCTATACCGTTGAC

- 10 Dabei sind auch kleinere Abweichungen (Degenerationen) der Erkennungssequenz umfasst, die dennoch eine Erkennung und Spaltung durch das jeweilige DSBI-Enzym ermöglichen. Derartige Abweichungen - auch in Zusammenhang mit unterschiedlichen Rahmenbedingungen wie beispielsweise Calcium oder Magnesium-
- 15 Konzentration - sind beschrieben (Argast GM et al. (1998) J Mol Biol 280:345-353). Ferner sind Kernsequenzen ("Core"-Sequenzen) dieser Erkennungssequenzen umfasst. Es ist bekannt, dass auch die inneren Anteile der Erkennungssequenzen für einen induzierten Doppelstrangbruch genügen und das die äußeren nicht unbedingt
- 20 relevant sind, jedoch die Effizienz der Spaltung mitbestimmen können. So kann beispielsweise für I-SceI eine 18bp-"Core"-Sequenz definiert werden.

- Besagte DSBI-Erkennungssequenzen können an verschiedenen
- 25 Positionen in oder in der Nähe eines Markerprotein-Gens lokalisiert werden und können - beispielsweise bei der Verwendung eines Transgens als Markerproteins - bereits bei der Konstruktion der Markerprotein-Expressionskassette eingebaut werden. Verschiedene Möglichkeiten der Lokalisation sind beispielhaft in den Fig. 2-A,
- 30 2-B, 3 und 5 sowie in den Beschreibungen dazu verdeutlicht.

- In einer weiteren vorteilhaften Ausführungsform umfasst die Insertionssequenz mindestens eine Homologiesequenz A, die eine ausreichende Länge und eine ausreichende Homologie zu einer
- 35 Sequenz A' in dem Markerproteingen aufweist, um eine homologe Rekombination zwischen A und A' zu gewährleisten. Bevorzugt ist Insertionssequenz von zwei Sequenzen A und B flankiert, die eine ausreichende Länge und eine ausreichende Homologie zu einer Sequenz A' bzw. B' in dem Markerproteingen aufweisen, um eine
- 40 homologe Rekombination zwischen A und A' bzw. B und B' zu gewährleisten.

- "Ausreichende Länge" meint in Bezug auf die Homologiesequenzen A, A' und B, B' bevorzugt Sequenzen von einer Länge von mindestens
- 45 100 Basenpaaren, bevorzugt mindestens 250 Basenpaaren, besonders bevorzugt von mindestens 500 Basenpaaren, ganz besonders bevor-



## 44

zugt von mindestens 1000 Basenpaaren, am meistens bevorzugt von mindestens 2500 Basenpaaren.

- "Ausreichende Homologie" meint in Bezug auf die Homologie-  
 5 sequenzen, bevorzugt Sequenzen die eine Homologie zueinander aufweisen von mindestens 70 %, bevorzugt 80 %, vorzugsweise mindestens 90 %, besonders bevorzugt mindestens 95 %, ganz besonders bevorzugt mindestens 99 %, am meisten bevorzugt 100% über eine  
 10 Länge von von mindestens 20 Basenpaaren, bevorzugt mindestens 50 Basenpaaren, besonders bevorzugt von mindestens 100 Basenpaaren, ganz besonders bevorzugt von mindestens 250 Basenpaaren, am meisten bevorzugt von mindestens 500 Basenpaaren.

- Unter Homologie zwischen zwei Nukleinsäuren wird die Identität  
 15 der Nukleinsäuresequenz über die jeweils gesamte Sequenzlänge verstanden, die durch Vergleich mit Hilfe des Programmalgorithmus GAP (Wisconsin Package Version 10.0, University of Wisconsin, Genetics Computer Group (GCG), Madison, USA) unter Einstellung folgender Parameter berechnet wird:

20

Gap Weight: 12

Length Weight: 4

Average Match: 2,912

Average Mismatch:-2,003

- 25 In einer weiteren bevorzugten Ausführungsform wird die Effizienz der Rekombination gesteigert durch Kombination mit Verfahren, die die homologe Rekombination fördern. Solche Verfahren sind beschrieben und umfassen beispielhaft die Expression von Proteinen wie RecA oder die Behandlung mit PARP-Inhibitoren. Es konnte  
 30 gezeigt werden, dass die intrachromosomale homologe Rekombination in Tabakpflanzen durch die Verwendung von PARP-Inhibitoren erhöht werden kann (Puchta H et al. (1995) Plant J 7:203-210). Durch den Einsatz dieser Inhibitoren kann die Rate der homologen Rekombination in den Rekombinationskonstrukten nach Induktion  
 35 des sequenzspezifischen DNA-Doppelstrangbruches und damit die Effizienz der Deletion der Transgensequenzen weiter erhöht werden. Verschiedene PARP Inhibitoren können dabei zum Einsatz kommen. Bevorzugt umfasst sind Inhibitoren wie 3-Aminobenzamid, 8-Hydroxy-2-methylquinazolin-4-on (NU1025), 1,11b-Dihydro-  
 40 [2H]benzopyrano[4,3,2-de]isoquinolin-3-on (GPI 6150), 5-Amino-isoquinolinon, 3,4-Dihydro-5-[4-(1-piperidinyloxy)]-1(2H)-isoquinolinon oder die in WO 00/26192, WO 00/29384, WO 00/32579, WO 00/64878, WO 00/68206, WO 00/67734, WO 01/23386 und WO 01/23390 beschriebenen Substanzen.

45

Weitere geeignete Methoden sind die Einführung von Nonsense-Mutationen in endogene Markerprotein Gene zum Beispiel mittels

## 45

Einführung von RNA/DNA-Oligonukleotiden in die Pflanze (Zhu et al. (2000) Nat Biotechnol 18(5):555-558). Punktmutationen können auch mittels DNA-RNA Hybriden erzeugt werden, die auch als "chimeraplasty" bekannt sind (Cole-Strauss et al. (1999) Nucl Acids Res 27(5):1323-1330; Kmiec (1999) Gene therapy American Scientist 87(3):240-247).

Die Methoden der dsRNAi, der Kosuppression mittels sense-RNA und der "VIGS" ("virus induced gene silencing") werden auch als "post-transcriptional gene silencing" (PTGS) bezeichnet. PTGS-Verfahren sind besonders vorteilhaft, weil die Anforderungen an die Homologie zwischen dem zu vermindernenden Markerprotein-Gen und der transgen exprimierten sense- oder dsRNA-Nukleinsäuresequenz geringer sind als beispielsweise bei einem klassischen antisense-Ansatz. So kann man unter Verwendung der Markerprotein-Nukleinsäuresequenzen aus einer Art auch die Expression von homologen Markerprotein-Proteinen in anderen Arten effektiv vermindern, ohne, dass die Isolierung und Strukturaufklärung der dort vorkommenden Markerprotein-Homologen zwingend erforderlich wäre. Dies erleichtert erheblich den Arbeitsaufwand.

"Einbringen" umfasst im Rahmen der Erfindung alle Verfahren, die dazu geeignet eine "anti-Markerprotein"-Verbindung, direkt oder indirekt, in eine Pflanze oder eine Zelle, Kompartiment, Gewebe, Organ oder Samen derselben einzuführen oder dort zu generieren. Direkte und indirekte Verfahren sind umfasst. Das Einbringen kann zu einer vorübergehenden (transienten) Präsenz einer "anti-Markerprotein"-Verbindung (beispielsweise einer dsRNA oder einer Rekombinase) führen oder aber auch zu einer dauerhaften (stabilen).

Gemäß der unterschiedlichen Natur der oben beschriebenen Ansätze kann die "anti-Markerprotein"-Verbindung ihre Funktion direkt (zum Beispiel durch Insertion in ein endogenes Markerprotein Gen) ausüben. Die Funktion kann aber auch indirekt nach Transkription in eine RNA (zum Beispiel bei antisense Ansätzen) oder nach Transkription und Translation in ein Protein (zum Beispiel bei Rekombinasen oder DSB-Enzymen) ausgeübt werden. Sowohl direkte als auch indirekt wirkende "anti-Markerprotein"-Verbindungen sind erfindungsgemäß umfasst.

Einführen umfasst beispielsweise Verfahren wie Transfektion, Transduktion oder Transformation.

"Anti-Markerprotein" Verbindungen umfasst somit beispielsweise auch Expressionskassetten, die eine Expression (d.h. Transkription und ggf. Translation) beispielsweise einer

MP-dsRNA, einer MP-antisenseRNA, einer sequenzspezifischen Rekombinase oder eines DSBI-Enzyms in einer pflanzlichen Zelle realisieren können.

- 5 "Expressionskassette" meint im Rahmen dieser Erfindung allgemein solche Konstruktionen in denen eine zu exprimierende Nukleinsäuresequenz in funktioneller Verknüpfung mit mindestens einer genetischen Kontrollsequenz - bevorzugt einer Promotorsequenz - steht. Expressionskassetten bestehen bevorzugt aus doppelsträngiger DNA und können eine lineare oder zirkuläre Struktur haben.

- Unter einer funktionellen Verknüpfung versteht man zum Beispiel die sequentielle Anordnung eines Promotors mit einer zu transkribierenden Nukleinsäuresequenz (beispielsweise kodierend für eines MP-dsRNA oder ein DSBI-Enzym) und ggf. weiterer regulativer Elemente wie zum Beispiel einem Terminator und/oder Polyadenylierungssignalen derart, dass jedes der regulativen Elemente seine Funktion bei der Transkription der Nukleinsäuresequenz, je nach Anordnung der Nukleinsäuresequenzen erfüllen kann. Funktion kann dabei beispielsweise die Kontrolle der Expression d.h. Transkription und/oder Translation der Nukleinsäuresequenz (beispielsweise kodierend für eine MP-dsRNA oder ein DSBI-Enzym) bedeuten. Kontrolle umfasst dabei beispielsweise das Initiieren, Steigerung, Steuerung oder Suppression der Expression d.h. Transkription und ggf. Translation. Dazu ist nicht unbedingt eine direkte Verknüpfung im chemischen Sinne erforderlich. Genetische Kontrollsequenzen, wie zum Beispiel Enhancer-Sequenzen, können ihre Funktion auch von weiter entfernten Positionen oder gar von anderen DNA-Molekülen aus auf die Zielsequenz ausüben. Bevorzugt sind Anordnungen, in denen die zu transkribierende Nukleinsäuresequenz hinter der als Promoter fungierenden Sequenz positioniert wird, so dass beide Sequenzen kovalent miteinander verbunden sind. Bevorzugt ist dabei der Abstand zwischen der Promotorsequenz und der transgen zu exprimierende Nukleinsäuresequenz geringer als 200 Basenpaare, besonders bevorzugt kleiner als 100 Basenpaare, ganz besonders bevorzugt kleiner als 50 Basenpaare.

- 40 Dem Fachmann sind verschiedene Wege bekannt, um zu einer der erfindungsgemässen Expressionskassette zu gelangen. Die Herstellung einer erfindungsgemässen Expressionskassette erfolgt beispielsweise bevorzugt durch direkte Fusion einer als Promoter fungierenden Nukleinsäuresequenz mit einer zu exprimierenden Nukleotidsequenz (beispielsweise kodierend für eines MP-dsRNA oder ein DSBI-Enzym). Die Herstellung einer funktionellen Verknüpfung kann mittels gängiger Rekombinations- und Klonierungs-

- techniken realisiert werden, wie sie beispielsweise in Maniatis T, Fritsch EF und Sambrook J (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY sowie in Silhavy TJ et al. (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY und in Ausubel FM et al. (1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience beschrieben sind.
- 10 Vorzugsweise umfassen die erfindungsgemässen Expressionskassetten 5'-stromaufwärts von der jeweiligen transgen zu exprimierenden Nukleinsäuresequenz einen Promotor und 3'-stromabwärts eine Terminatorsequenz als zusätzliche genetische Kontrollsequenz, sowie gegebenenfalls weitere übliche regulative Elemente, und zwar
- 15 jeweils funktionell verknüpft mit der transgen zu exprimierenden Nukleinsäuresequenz.

- Der Begriff der "genetischen Kontrollsequenzen" ist breit zu verstehen und meint all solche Sequenzen, die einen Einfluss
- 20 auf das Zustandekommen oder die Funktion der erfindungsgemässen Expressionskassette haben. Genetische Kontrollsequenzen gewährleisten zum Beispiel die Transkription und gegebenenfalls Translation in prokaryotischen oder eukaryotischen Organismen. Genetische Kontrollsequenzen sind beispielsweise beschrieben bei
- 25 "Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990)" oder "Gruber and Crosby, in: Methods in Plant Molecular Biology and Biotechnology, CRC Press, Boca Raton, Florida, eds.: Glick and Thompson, Chapter 7, 89-108" sowie den dort aufgewiesenen Zitaten.
- 30 Genetische Kontrollsequenzen umfassen insbesondere in Pflanzen funktionelle Promotoren. Als bevorzugte Promotoren für die Expressionskassetten ist grundsätzlich jeder Promotor geeignet, der die Expression von Genen, insbesondere Fremdgenen, in
- 35 Pflanzen steuern kann.

- Pflanzenspezifische oder in Pflanzen bzw. pflanzlichen Zelle funktionelle Promotoren meint grundsätzlich jeden Promotor, der die Expression von Genen, insbesondere Fremdgenen, in mindestens
- 40 einer Pflanze oder einem Pflanzenteil, -zelle, -gewebe, -kultur steuern kann. Dabei kann die Expression beispielsweise konstitutiv, induzierbar oder entwicklungsabhängig sein. Bevorzugt sind:

## a) Konstitutive Promotoren

5 "Konstitutive" Promotoren meint solche Promotoren, die eine Expression in zahlreichen, bevorzugt allen, Geweben über einen größeren Zeitraum der Pflanzenentwicklung, bevorzugt zu allen Zeitpunkten der Pflanzenentwicklung, gewährleisten (Benfey et al. (1989) EMBO J 8:2195-2202). Vorzugsweise verwendet man insbesondere einen pflanzlichen Promotor oder einen Promotor, der einem Pflanzenvirus entstammt. Insbesondere bevorzugt ist der Promotor des 35S-Transkriptes des CaMV Blumenkohlmosaikvirus (Franck et al. (1980) Cell 21:285-294; Odell et al. (1985) Nature 313:810-812; Shewmaker et al. (1985) Virology 140:281-288; Gardner et al. (1986) Plant Mol Biol 6:221-228) oder der 19S CaMV Promotor (US 5,352,605; WO 84/02913; Benfey et al. (1989) EMBO J 8:2195-2202) sowie der Promotor des Nitrilase-1 Gens aus Arabidopsis thaliana (GenBank Acc.-No.: Y07648, Nukleotide 2456 (alternativ 2861) bis 4308 oder alternativ 4340 oder 4344. (beispielsweise bp 2456 bis 4340)).

20 Ein weiterer geeigneter konstitutiver Promotor ist der "Rubisco small subunit (SSU)"-Promotor (US 4,962,028), der LeguminB-Promotor (GenBank Acc.-No.: X03677), der Promotor der Nopalinsynthase aus Agrobacterium, der TR-Doppelpromotor, 25 der OCS (Octopin Synthase) Promotor aus Agrobacterium, der Ubiquitin Promotor (Holtorf S et al. (1995) Plant Mol Biol 29:637-649), der Ubiquitin 1 Promotor (Christensen et al. (1992) Plant Mol Biol 18:675-689; Bruce et al. (1989) Proc Natl Acad Sci USA 86:9692-9696), der Smas Promotor, der 30 Cinnamylalkoholdehydrogenase-Promotor (US 5,683,439), die Promotoren der vakuolärer ATPase Untereinheiten oder der Promotor eines prolinreichen Proteins aus Weizen (WO 91/13991), sowie weitere Promotoren von Genen, deren konstitutive Expression in Pflanzen dem Fachmann bekannt ist.

35

## b) Gewebespezifische Promotoren

Bevorzugt sind Promotoren mit Spezifitäten für die Antheren, Ovarien, Blüten, Blätter, Stengel, Wurzeln oder Samen.

40

Samenspezifische Promotoren umfassen zum Beispiel den Promotor des Phaseolins (US 5,504,200; Bustos MM et al. (1989) Plant Cell 1(9):839-53), des 2S Albumins (Joseffson LG et al. (1987) J Biol Chem 262:12196-12201), des 45 Legumins (Shirsat A et al. (1989) Mol Gen Genet 215(2):326-331), des USP (unknown seed protein; Bäumlein H et al. (1991) Mol Gen Genet 225(3):459-67), des Napins

- (US 5,608,152; Stalberg K et al. (1996) L Planta 199:515-519), des Saccharosebindeproteins (WO 00/26388), des Legumins B4 (LeB4; Baumlein H et al. (1991) Mol Gen Genet 225: 121-128; Baumlein et al. (1992) Plant Journal 2(2):233-9; Fiedler U et al. (1995) Biotechnology (NY) 13(10):1090f), des Oleosins (WO 98/45461) oder des Bce4 (WO 91/13980). Weitere geeignete samenspezifische Promotoren sind die der Gene kodierend für das "High Molecular Weight Glutenin" (HMWG), Gliadin, Verzweigungsenzym, ADP Glucose Pyrophosphatase (AGPase) oder Stärkesynthase. Bevorzugt sind ferner Promotoren, die eine samenspezifische Expression in Monokotyledonen wie Mais, Gerste, Weizen, Roggen, Reis etc. erlauben. Vorteilhaft eingesetzt werden können der Promoter des lpt2- oder lpt1-Gens (WO 95/15389, WO 95/23230) oder die Promotoren beschrieben in WO 99/16890 (Promotoren des Hordeins, Glutelins, Oryzins, Prolamins, Gliadins, Zeins, Kasirins oder Secalins). Weitere samenspezifische Promotoren sind beschrieben in WO 89/03887.
- Knollen-, Speicherwurzel- oder Wurzel-spezifische Promotoren umfassen beispielsweise den Patatin Promotor Klasse I (B33) oder den Promotor des Cathepsin D Inhibitors aus Kartoffel.
- Blattspezifische Promotoren umfassen beispielsweise den Promotor der cytosolischen FBPase aus Kartoffel (WO 97/05900), den SSU Promotor (small subunit) der Rubisco (Ribulose-1,5-bisphosphatcarboxylase) oder den ST-LSI Promotor aus Kartoffel (Stockhaus et al. (1989) EMBO J 8:2445-2451).
- Blütenspezifische Promotoren umfassen beispielsweise den Phytoen Synthase Promotor (WO 92/16635) oder den Promotor des P-rr Gens (WO 98/22593).
- Antheren-spezifische Promotoren umfassen beispielsweise den 5126-Promotor (US 5,689,049, US 5,689,051), den glob-1 Promotor und den  $\gamma$ -Zein Promotor.
- c) Chemisch induzierbare Promotoren
- Chemisch induzierbare Promotor erlauben es, die Expression abhängig von einem exogenen Stimulus zu steuern (Übersichtsartikel: Gatz et al. (1997) Ann Rev Plant Physiol Plant Mol Biol 48:89-108). Beispielhaft seien zu nennen: Der PRP1 Promotor (Ward et al. (1993) Plant Mol Biol 22:361-366),

## 50

ein durch Salicylsäure induzierbarer Promotor (WO 95/19443), ein durch Benzolsulfonamid-induzierbarer Promotor (EP-A 0 388 186), ein durch Tetrazyklin-induzierbarer Promotor (Gatz et al. (1992) Plant J 2:397-404), ein durch Abscisinsäure induzierbarer Promotor (EP 0 335 528) bzw. ein durch Ethanol- oder Cyclohexanon-induzierbarer Promotor (WO 93/21334). Ferner geeignet ist der Promotor des Glutathione-S-transferase Isoform II Gens (GST-II-27), der durch exogen applizierte Safener wie z.B. N,N-Diallyl-2,2-dichloroacetamid aktiviert werden kann (WO 93/01294) und in zahlreichen Geweben von sowohl Monokotyledonen als auch Dikotyledonen funktionell ist.

Besonders bevorzugt sind konstitutive oder induzierbare Promotoren.

Ferner sind für die gezielte Expression in den Plastiden plastiden-spezifische Promotoren bevorzugt. Geeignete Promotoren sind beispielsweise beschrieben in WO 98/55595 oder WO 97/06250. Zu nennen sind das rpo B Promotorelement, das atoB Promotorelement, das clpP Promotorelement (siehe auch WO 99/46394) oder das 16SrDNA Promotorelement. Weiterhin sind virale Promotoren geeignet (WO 95/16783).

Eine gezielte plastidäre Expression kann auch erreicht werden, wenn man zum Beispiel einen bakteriellen oder Bakteriophagen Promotor verwendet, die resultierende Expressionskassette in die plastidäre DNA einbringt und die Expression dann durch ein Fusionsprotein aus einer bakteriellen oder Bakteriophagen Polymerase und einem plastidären Transitpeptid exprimiert. Ein entsprechendes Verfahren ist in US 5,925,806 beschrieben.

Genetische Kontrollsequenzen umfassen ferner auch die 5'-untranslatierte Regionen, Introns oder nichtkodierende 3'-Region von Genen wie beispielsweise das Actin-1 Intron, oder die Adh1-S Introns 1, 2 und 6 (allgemein: The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994)). Es ist gezeigt worden, dass diese eine signifikante Funktionen bei der Regulation der Genexpression spielen können. So wurde gezeigt, dass 5'-untranslatierte Sequenzen die transiente Expression heterologer Gene verstärken können. Sie können ferner die Gewebespezifität fördern (Rouster J et al. (1998) Plant J. 15:435-440). Beispielhaft für Translationsverstärker sei die 5'-Leadersequenz aus dem Tabak-Mosaik-Virus zu nennen (Gallie et al. (1987) Nucl Acids Res 15:8693-8711).

## 51

Als Kontrollsequenzen geeignete Polyadenylierungssignale sind insbesondere Polyadenylierungssignale pflanzlicher Gene sowie T-DNA Polyadenylierungssignale aus *Agrobacterium tumefaciens*. Beispiele für besonders geeignete Terminatorsequenzen sind

- 5 der OCS (Octopin-Synthase)-Terminator und der NOS (Nopalin-Synthase)-Terminator (Depicker A et al (1982) J Mol Appl Genet 1:561-573), als auch die Terminatoren von Sojabohnen Actin, RUBISCO oder alpha-Amylase aus Weizen (Baulcombe DC et al (1987) Mol Gen Genet 209:33-40).

10

Die Expressionskassette kann vorteilhafterweise eine oder mehrere sogenannte "enhancer Sequenzen" funktionell verknüpft mit dem Promoter enthalten, die eine erhöhte transgene Expression der Nukleinsäuresequenz ermöglichen.

15

- Genetische Kontrollsequenzen meint ferner Sequenzen, die für Fusionsproteine bestehend aus einer Signalpeptidsequenz kodieren. Die Expression eines Zielgenes ist in jedem gewünschten Zellkompartiment, wie z.B. dem Endomembransystem, der Vakuole und
- 20 den Chloroplasten möglich. Durch Nutzung des sekretorischen Weges sind gewünschte Glykosylierungsreaktionen, besondere Faltungen u.ä. möglich. Auch die Sekretion des Zielproteins zur Zelloberfläche bzw. die Sezernierung ins Kulturmedium, beispielsweise bei Nutzung suspensionskultivierter Zellen oder Protoplasten ist
- 25 möglich. Die dafür notwendigen Targetsequenzen können sowohl in einzelnen Vektorvariationen berücksichtigt werden als auch durch Verwendung einer geeigneten Klonierungsstrategie gemeinsam mit dem zu klonierenden Zielgen in den Vektor mit eingebracht werden. Als Targetsequenzen können sowohl Gen-eigene, sofern vorhanden,
- 30 oder heterologe Sequenzen genutzt werden. Zusätzliche, heterologe zur funktionellen Verknüpfung bevorzugte aber nicht darauf beschränkte Sequenzen sind weitere Targeting-Sequenzen zur Gewährleistung der subzellulären Lokalisation im Apoplasten, in der Vakuole, in Plastiden, im Mitochondrium, im Endoplas-
- 35 matischen Retikulum (ER), im Zellkern, in Ölkörperchen oder anderen Kompartimenten; sowie Translationsverstärker wie die 5'-Leadersequenz aus dem Tabak-Mosaik-Virus (Gallie et al. (1987) Nucl Acids Res 15: 8693-8711) und dergleichen. Das Verfahren, an sich nicht in den Plastiden lokalisierte Proteine, gezielt
- 40 in die Plastiden zu transportieren ist beschrieben (Klosgen RB und Weil JH (1991) Mol Gen Genet 225(2):297-304; Van Breusegem F et al. (1998) Plant Mol Biol 38(3):491-496).

- Als Kontrollsequenzen sind weiterhin solche zu verstehen, die
- 45 eine homologe Rekombination bzw. Insertion in das Genom eines Wirtsorganismus ermöglichen oder die Entfernung aus dem Genom erlauben. Methoden wie die cre/lox-Technologie erlauben eine



## 52

gewebsspezifische, unter Umständen induzierbare Entfernung der Expressionskassette aus dem Genom des Wirtsorganismus (Sauer B. Methods. 1998; 14(4):381-92). Hier werden bestimmte flankierende Sequenzen dem Zielgen angefügt (lox-Sequenzen), die später eine  
5 Entfernung mittels der cre-Rekombinase ermöglichen.

Bevorzugt kann die Expressionskassette, bestehend aus einer Verknüpfung von Promoter und zu transkribierender Nukleinsäuresequenz, integriert in einem Vektor vorliegen und durch zum Bei-  
10 spiel Transformation - nach einem der unten beschriebenen Verfahren - in die pflanzliche Zelle oder Organismus eingebracht werden.

"Transgen" meint bevorzugt - beispielsweise in Bezug auf eine  
15 transgene Expressionskassette, einen transgenen Expressionsvektor, einen transgenen Organismus oder Verfahren zur transgenen Expression von Nukleinsäuren - alle solche durch gentechnische Methoden zustande gekommene Konstruktionen oder Verfahren unter Verwendung derselben, in denen entweder

20

a) die zu exprimierende Nukleinsäuresequenz, oder

b) der mit der zu exprimierenden Nukleinsäuresequenz gemäß a) funktionell verknüpfte Promotor, oder

25

c) (a) und (b)

sich nicht in ihrer natürlichen, genetischen Umgebung (d.h. an ihrem natürlichen chromosomalen Locus) befinden oder durch gen-  
30 technische Methoden modifiziert wurden, wobei die Modifikation beispielhaft eine Substitution, Addition, Deletion, Inversion oder Insertion eines oder mehrerer Nukleotidreste sein kann. Natürliche genetische Umgebung meint den natürlichen chromosomalen Locus in dem Herkunftsorganismus oder das Vorliegen  
35 in einer genomischen Bibliothek.

"Transgen" meint in Bezug auf eine Expression ("transgene Expression") bevorzugt all solche unter Einsatz einer transgenen Expressionskassette, transgenen Expressionsvektor oder transgenen  
40 Organismus - entsprechend dem oben gegebenen Definitionen - realisierten Expressionen.

Die im Rahmen des erfindungsgemässen Verfahren zum Einsatz kommenden DNA-Konstrukte und die von ihnen abgeleiteten  
45 Vektoren können weitere Funktionselemente enthalten. Der Begriff Funktionselement ist breit zu verstehen und meint all solche Elemente, die einen Einfluss auf Herstellung, Vermehrung

oder Funktion der DNA-Konstrukte oder von diesen abgeleitete Vektoren oder Organismen haben. Beispielhaft aber nicht einschränkend seien zu nennen:

#### 5 1. Selektionsmarker

Selektionsmarker umfassen beispielsweise solche Nukleinsäure- oder Proteinsequenzen, deren Expression einer Zelle, Gewebe oder Organismus einen Vorteil (positiver Selektionsmarker) oder Nach-  
10 teil (negativer Selektionsmarker) gegenüber Zellen vermittelt, die diese Nukleinsäure oder Protein nicht exprimieren. Positive Selektionsmarker wirken beispielsweise dadurch, dass eine auf die Zelle inhibitorisch wirkende Substanz detoxifiziert wird (Bsp. Antibiotika-/Herbizidresistenz), oder eine Substanz  
15 gebildet wird, welche der Pflanze unter den gewählten Bedingungen verbesserte Regeneration oder erhöhtes Wachstum ermöglicht (zum Beispiel nutritive Marker, hormonproduzierender Marker wie ipt; s.u.). Eine andere Form positiver Selektionsmarker umfasst mutierte Proteine oder RNAs, die gegenüber einem selektiven Agens  
20 nicht empfindlich sind (beispielsweise 16S rRNA Mutanten, die unempfindlich gegenüber Spectinomycin sind). Negative Selektionsmarker wirken beispielsweise dadurch, dass sie die Bildung einer toxischen Substanz in den transformierten Zellen katalysieren (zum Beispiel das codA Gen).

25

##### 1.1 Positive Selektionsmarker:

Die DNA-Konstrukte können zur weiteren Steigerung der Effizienz zusätzliche positive Selektionsmarker umfassen. In einer bevor-  
30 zugten Ausführungsform kann so das erfindungsgemäße Verfahren im Form einer doppelten Selektion realisiert werden, wobei zusammen mit der zu insertierenden Nukleinsäuresequenz ein Sequenz kodierend für eine Resistenz gegen mindestens ein Toxin, Antibiotikum oder Herbizid eingebracht wird, und die Selektion  
35 zusätzlich unter Einsatz des Toxins, Antibiotikums oder Herbizids erfolgt.

Entsprechende Proteine und Sequenzen von positiven Selektionsmarkern sowie Selektionsverfahren sind dem Fachmann geläufig. Der  
40 Selektionsmarker verleiht den erfolgreich transformierten Zellen eine Resistenz gegen ein Biozid (zum Beispiel ein Herbizid wie Phosphinothricin, Glyphosat oder Bromoxynil), einen Metabolismusinhibitor wie 2-Desoxyglucose-6-phosphat (WO 98/45456) oder ein Antibiotikum, wie zum Beispiel Tetracycline, Ampicillin, Kana-  
45 mycin, G 418, Neomycin, Bleomycin oder Hygromycin. Beispielhaft als Selektionsmarker seien genannt:

## 54

- Phosphinothricinacetyltransferasen (PAT), welche die freie Aminogruppe des Glutaminsynthaseinhibitors Phosphinothricin (PPT) acetylieren und damit eine Detoxifizierung des PPT erreichen (de Block et al. (1987) EMBO J 6:2513-2518) (auch Bialophos® Resistenzgen (bar) genannt). Entsprechende Sequenzen sind dem Fachmann bekannt (aus *Streptomyces hygrosopicus* GenBank Acc.-No.: X17220 und X05822, aus *Streptomyces viridochromogenes* GenBank Acc.-No.: M 22827 und X65195; US 5,489,520). Ferner sind synthetische Gene für die Expression in Plastiden beschrieben. Ein synthetisches PAT-Gen ist beschrieben in Becker et al. (1994) Plant J 5:299-307. Die Gene verleihen Resistenz gegen das Herbizid Bialaphos oder Glufosinat und sind vielbenutzer Marker in transgenen Pflanzen (Vickers JE et al. (1996) Plant Mol Biol Reporter 14:363-368; Thompson CJ et al. (1987) EMBO J 6:2519-2523).
- 5-Enolpyruvylshikimat-3-phosphatsynthasen (EPSPS), die eine Resistenz gegen Glyphosat (N-(phosphonomethyl)glycin) verleihen. Das unselektive Herbizid Glyphosat hat die 5-Enolpyruvyl-3-phosphoshikimatsynthase (EPSPS) als molekulares Target. Diese hat eine Schlüsselfunktion in der Biosynthese aromatischer Aminosäuren in Mikroben und Pflanzen, jedoch nicht in Säugern (Steinrücken HC et al. (1980) Biochem Biophys Res Commun 94:1207-1212; Levin JG und Sprinson DB (1964) J Biol Chem 239:1142-1150; Cole DJ (1985) Mode of action of glyphosate a literature analysis, p. 48-74. In: Grossbard E und Atkinson D (eds.). The herbicide glyphosate. Butterworths, Boston.). Glyphosat-tolerante EPSPS Varianten werden bevorzugt als Selektionsmarker verwendet (Padgett SR et al. (1996). New weed control opportunities: development of soybeans with a Roundup Ready™ gene. In: Herbicide Resistant Crops (Duke, S.O., ed.), pp. 53-84. CRC Press, Boca Raton, FL; Saroha MK und Malik VS (1998) J Plant Biochemistry and Biotechnology 7:65-72). Das EPSPS Gen des Agrobakterium sp. strain CP4 hat eine natürliche Toleranz gegen Glyphosat, die auf entsprechende transgene Pflanzen transferiert werden kann. Das CP4 EPSPS Gen wurde aus Agrobakterium sp. strain CP4 kloniert (Padgett SR et al. (1995) Crop Science 35(5):1451-1461). Sequenzen von EPSPS-Enzymen, die Glyphosat-tolerant sind, sind beschrieben (u.a. in US 5,510,471; US 5,776,760; US 5,864,425; US 5,633,435; US 5,627,061; US 5,463,175; EP 0 218 571). Weitere Sequenzen sind beschrieben unter GenBank Acc.-No: X63374 oder M10947.

## 55

- 5        -    Glyphosat® degradierende Enzyme (gox Gen; Glyphosatoxido-reduktase). GOX (beispielsweise die Glyphosatoxidoreductase aus *Achromobacter* sp.) katalysiert die Spaltung einer C-N Bindung im Glyphosat, welches so zu Aminomethylphosphonsäure (AMPA) und Glyoxylat umgesetzt wird. GOX kann dadurch eine Resistenz gegen Glyphosat vermitteln (Padgett et al. (1996) *J Nutr* 126(3):702-16; Shah D et al. (1986) *Science* 233:478-481).
- 10    -    Das deh Gen kodiert für eine Dehalogenase, die Dalapon® inaktiviert (GenBank Acc.-No.: AX022822, AX022820 sowie WO 99/27116)
- 15    -    Die bxn Gene kodieren für Bromoxynil degradierende Nitrilaseenzyme (Genbank Acc.-No: E01313 und J03196).
- 20    -    Neomycinphosphotransferasen verleihen eine Resistenz gegen Antibiotika (Aminoglykoside) wie Neomycin, G418, Hygromycin, Paromomycin oder Kanamycin, indem sie durch eine Phosphorylierungsreaktion deren inhibierende Wirkung reduzieren. Besonders bevorzugt ist das nptII Gen. Sequenzen können aus der GenBank erhalten werden (AF080390; AF080389). Zudem ist das Gen bereits Bestandteil zahlreicher Expressionsvektoren und kann unter Verwendung von dem Fachmann geläufigen Verfahren aus diesen isoliert werden (AF234316; AF234315; AF234314). Das NPTII Gen kodiert für eine Aminoglycosid-3'-O-phosphotransferase aus *E.coli*, Tn5 (GenBank Acc.-No: U00004 Position 1401-2300; Beck et al. (1982) *Gene* 19 327-336).
- 25    -    Das DOG<sup>R</sup>1-Gen wurde aus der Hefe *Saccharomyces cerevisiae* isoliert (EP-A 0 807 836) und kodiert für eine 2-Desoxyglukose-6-phosphat Phosphatase, die eine Resistenz gegenüber 2-DOG verleiht (Randez-Gil et al. (1995) *Yeast* 11:1233-1240; Sanz et al. (1994) *Yeast* 10:1195-1202, GenBank Acc.-No.: NC001140; Position 194799-194056).
- 30    -    Acetolactatsynthasen, die eine Resistenz gegen Imidazolinon/Sulfonylurea-Herbizide verleihen (GenBank Acc.-No.: X51514; Sathasivan K et al. (1990) *Nucleic Acids Res.* 18(8):2188; AB049823; AF094326; X07645; X07644; A19547; A19546; A19545; I05376; I05373; AL133315)
- 35    -    Hygromycinphosphotransferasen (z.B. GenBank Acc.-No.: X74325) die eine Resistenz gegen das Antibiotikum Hygromycin verleihen. Das Gen ist Bestandteil zahlreicher Expressionsvektoren und kann unter Verwendung von dem Fachmann
- 40    -    Hygromycinphosphotransferasen (z.B. GenBank Acc.-No.: X74325) die eine Resistenz gegen das Antibiotikum Hygromycin verleihen. Das Gen ist Bestandteil zahlreicher Expressionsvektoren und kann unter Verwendung von dem Fachmann
- 45    -    Hygromycinphosphotransferasen (z.B. GenBank Acc.-No.: X74325) die eine Resistenz gegen das Antibiotikum Hygromycin verleihen. Das Gen ist Bestandteil zahlreicher Expressionsvektoren und kann unter Verwendung von dem Fachmann

## 56

geläufigen Verfahren (wie beispielsweise Polymerasekettenreaktion) aus diesen isoliert werden (GenBank Acc-No.: AF294981; AF234301; AF234300; AF234299; AF234298; AF354046; AF354045)

5

- Resistenzgene gegen

- a) Chloramphenicol (Chloramphenicolacetyltransferase),
- 10 b) Tetracyclin (u.a. GenBank Acc-No.: X65876; X51366). Zudem ist das Gen bereits Bestandteil zahlreicher Expressionsvektoren und kann unter Verwendung von dem Fachmann geläufigen Verfahren (wie beispielsweise Polymerasekettenreaktion) aus diesen isoliert werden
- 15 c) Streptomycin (u.a. GenBank Acc.-No.: AJ278607).
- d) Zeocin, das entsprechende Resistenzgen ist Bestandteil zahlreicher Klonierungsvektoren (z.B. GenBank Acc.-No.: L36849) und kann unter Verwendung von dem Fachmann geläufigen Verfahren (wie beispielsweise Polymerasekettenreaktion) aus diesen isoliert werden.
- 20 e) Ampicillin ( $\beta$ -Lactamase Gen; Datta N, Richmond MH (1966) Biochem J 98(1):204-9; Heffron F et al (1975) J. Bacteriol 122: 250-256; Bolivar F et al. (1977) Gene 2:95-114). Die Sequenz ist Bestandteil zahlreicher Klonierungsvektoren und kann unter Verwendung von dem Fachmann geläufigen Verfahren (wie beispielsweise Polymerasekettenreaktion) aus diesen isoliert werden.
- 25
- 30

Gene wie die Isopentenyltransferase aus Agrobacterium tumefaciens (strain:PO22) (Genbank Acc.-No.: AB025109) können auch als Selektionsmarker eingesetzt werden. Das ipt Gen ist ein Schlüsselenzym der Cytokinin-Biosynthese. Seine Überexpression erleichtert die  
 35 Regeneration von Pflanzen (z.B. Selektion auf Cytokinin-freiem Medium). Das Verfahren zur Nutzung des ipt Gens ist beschrieben (Ebinuma H et al. (2000) Proc Natl Acad Sci USA 94:2117-2121; Ebinuma H et al. (2000) Selection of Marker-free transgenic  
 40 plants using the oncogenes (ipt, rol A, B, C) of Agrobacterium as selectable markers, In Molecular Biology of Woody Plants. Kluwer Academic Publishers).

Verschiedene weitere positive Selektionsmarker, die den transformierten Pflanzen einen Wachstumsvorteil gegenüber nicht-transformierten verleihen, sowie Verfahren zu ihrer Verwendung sind  
 45 u.a. beschrieben in EP-A 0 601 092. Beispielhaft sind zu nennen

## 57

$\beta$ -Glucuronidase (in Verbindung mit z.B. Cytokininglucuronid), Mannose-6-phosphat-Isomerase (in Verbindung mit Mannose), UDP-Galaktose-4-Epimerase (in Verbindung mit z.B. Galactose).

- 5 Für einen in Plastiden funktionellen Selektionsmarker sind insbesondere solche bevorzugt, die eine Resistenz gegen Spectinomycin, Streptomycin, Kanamycin, Lincomycin, Gentamycin, Hygromycin, Methotrexat, Bleomycin, Phleomycin, Blastocidin, Sulfonamid,, Phosphinotricin, Chlorsulfuron, Bromoxymil, Glyphosat, 10 2,4-Datrazin, 4-methyltryptophan, Nitrat, S-aminoethyl-L-cysteine, Lysin/Threonin, Aminoethyl-Cystein oder Betainaldehyd verleihen. Besonders bevorzugt sind die Gene aadA, nptII, BADH, FLARE-S (eine Fusion aus aadA und GFP, beschrieben bei Khan MS & Maliga P (1999) Nature Biotech 17:910-915). Geeignet ist vor 15 allem das aadA Gen (Svab Z und Maliga P (1993) Proc Natl Acad Sci USA 90:913-917). Ferner beschrieben sind modifizierte 16S rDNA sowie die Betainealdehyddehydrogenase (BADH) aus Spinat (Daniell H et al. (2001) Trends Plant Science 6:237-239; Daniell H et al. (2001) Curr Genet 39:109-116; WO 01/64023; WO 01/64024; 20 WO 01/64850). Auch lethal wirkende Agenzen wie beispielsweise Glyphosat können in Verbindung mit entsprechend detoxifizierenden oder resistenten Enzymen genutzt werden (WO 01/81605).

- Die jeweils für die Selektion verwendeten Konzentrationen 25 der Antibiotika, Herbizide, Biozide oder Toxine müssen an die jeweiligen Testbedingungen bzw. Organismen angepasst werden. Beispielfhaft seien für Pflanzen zu nennen Kanamycin (Km) 50 mg/L, Hygromycin B 40 mg/L, Phosphinothricin (Ppt) 6 mg/L, Spectinomycin (Spec) 500 mg/L.

30

## 2. Reportergene

- Reportergene kodieren für leicht quantifizierbare Proteine und gewährleisten so über Eigenfarbe oder Enzymaktivität eine 35 Bewertung der Transformationseffizienz, des Expressionsortes oder -zeitpunktes. Ganz besonders bevorzugt sind dabei Gene kodierend für Reporter-Proteine (siehe auch Schenborn E, Groskreutz D (1999) Mol Biotechnol 13(1):29-44) wie

- 40 - "green fluorescence protein" (GFP) (Chui WL et al. (1996) Curr Biol 6:325-330; Leffell SM et al. (1997) Biotechniques 23(5):912-8; Sheen et al. (1995) Plant J 8(5):777-784; Haseloff et al. (1997) Proc Natl Acad Sci USA 94(6): 2122-2127; Reichel et al. (1996) Proc Natl Acad Sci USA 45 93(12):5888-5893; Tian et al. (1997) Plant Cell Rep 16:267-271; WO 97/41228)

- Chloramphenicoltransferase
- Luziferase (Millar et al. (1992) Plant Mol Biol Rep 10: 324-414; Ow et al. (1986) Science 234:856-859); erlaubt Biolumineszenzdetektion
- $\beta$ -Galactosidase (kodiert für ein Enzym für das verschiedenen chromogene Substrate zur Verfügung stehen)
- 10 -  $\beta$ -Glucuronidase (GUS) (Jefferson et al. (1987) EMBO J 6: 3901-3907) oder das uidA Gen (kodieren für Enzyme für die verschiedene chromogene Substrate zur Verfügung stehen)
- R-Locus Genprodukt, das die Produktion von Anthocyanin-pigmenten (rote Färbung) in pflanzlichen Gewebe reguliert und so eine direkte Analyse der Promoteraktivität ohne Zugabe zusätzlicher Hilfsstoffe oder chromogener Substrate ermöglicht (Dellaporta et al. (1988) In: Chromosome Structure and Function: Impact of New Concepts, 18<sup>th</sup> Stadler Genetics Symposium, 11:263-282)
- 15 - Tyrosinase (Katz et al. (1983) J Gen Microbiol 129:2703-2714), Enzym, das Tyrosin zu DOPA und Dopaquinon oxidiert, die infolge das leicht nachweisbare Melanin bilden.
- 25 - Aequorin (Prasher et al. (1985) Biochem Biophys Res Commun 126(3):1259-1268), kann in der Calcium-sensitiven Biolumineszenzdetektion verwendet werden.
- 30 3. Replikationsursprünge, die eine Vermehrung der erfindungsgemäßen Expressionskassetten oder Vektoren in zum Beispiel E.coli gewährleisten. Beispielhaft seien genannt ORI (origin of DNA replication), der pBR322 ori oder der P15A ori (Sambrook et al.: Molecular Cloning. A Laboratory Manual, 2<sup>nd</sup> ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).
- 35 4. Elemente zum Beispiel "Bordersequenzen", die einen Agrobakterien-vermittelte Transfer in Pflanzenzellen für die Übertragung und Integration ins Pflanzengenom ermöglichen, wie zum Beispiel die rechte oder linke Begrenzung der T-DNA oder die vir-Region.
- 40 5. Multiple Klonierungsregionen (MCS) erlauben und erleichtern die Insertion eines oder mehrerer Nukleinsäuresequenzen.
- 45

Die Einführung von Nukleinsäuresequenzen (z.B. Expressionskassetten) in einen pflanzlichen Organismus oder Zellen, Geweben, Organe, Teile bzw. Samen desselben kann vorteilhaft unter Verwendung von Vektoren realisiert werden, in denen diese Sequenzen  
5 enthalten sind. Vektoren können beispielhaft Plasmide, Cosmide, Phagen, Viren oder auch Agrobakterien sein. Die Sequenzen können in den Vektor (bevorzugt ein Plasmidvektor) über geeignete Restriktionsschnittstellen insertiert werden. Der entstandene Vektor kann zunächst in *E.coli* eingeführt und amplifiziert  
10 werden. Korrekt transformierte *E.coli* werden selektioniert, gezüchtet und der rekombinante Vektor mit dem Fachmann geläufigen Methoden gewonnen. Restriktionsanalyse und Sequenzierung können dazu dienen, den Klonierungsschritt zu überprüfen. Bevorzugt sind solche Vektoren, die eine stabile Integration in das Wirtsgenom  
15 ermöglichen.

Die Herstellung eines transformierten Organismus (bzw. einer transformierten Zelle oder Gewebes) erfordert, dass die entsprechende DNA (z.B. der Transformationsvektor) oder RNA in die  
20 entsprechende Wirtszelle eingebracht wird. Für diesen Vorgang, der als Transformation (oder Transduktion bzw. Transfektion) bezeichnet wird, steht eine Vielzahl von Methoden und Vektoren zur Verfügung (Keown et al. (1990) *Methods in Enzymology* 185:527-537; *Plant Molecular Biology and Biotechnology* (CRC  
25 Press, Boca Raton, Florida), Kapitel 6/7, S. 71-119 (1993); White FF (1993) *Vectors for Gene Transfer in Higher Plants*; in: *Transgenic Plants*, Bd. 1, Engineering and Utilization, Hrsgb.: Kung und Wu R, Academic Press, 15-38; Jenes B et al. (1993) *Techniques for Gene Transfer*, in: *Transgenic Plants*, Bd. 1,  
30 Engineering and Utilization, Hrsgb.: Kung und R. Wu, Academic Press, S.128-143; Potrykus (1991) *Annu Rev Plant Physiol Plant Molec Biol* 42:205-225; Halford NG, Shewry PR (2000) *Br Med Bull* 56(1):62-73).

35 Beispielsweise kann die DNA oder RNA direkt durch Mikroinjektion (WO 92/09696, WO 94/00583, EP-A 0 331 083, EP-A 0 175 966) oder durch Bombardierung mit DNA bzw. RNA-beschichteten Mikropartikeln (biolistische Verfahren mit der Genkanone "particle bombardment"; US 5,100,792; EP-A 0 444 882; EP-A 0 434 616; Fromm ME et al.  
40 (1990) *Bio/Technology* 8(9):833-9; Gordon-Kamm et al. (1990) *Plant Cell* 2:603) eingeführt werden. Auch kann die Zelle chemisch, zum Beispiel mit Polyethylenglycol, permeabilisiert werden, so dass die DNA durch Diffusion in die Zelle gelangen kann. Die DNA kann auch durch Protoplastenfusion mit anderen DNA-enthaltenden Ein-  
45 heiten wie Minicells, Zellen, Lysosomen oder Liposomen (Freeman et al. (1984) *Plant Cell Physiol.* 29:1353ff; US 4,536,475) erfolgen. Elektroporation ist eine weitere geeignete Methode



- zur Einführung von DNA, bei der die Zellen reversibel durch einen elektrischen Impuls permeabilisiert werden (EP-A 290 395, WO 87/06614). Weitere Verfahren umfassen die Calciumphosphat-vermittelte Transformation, die DEAE-Dextran-vermittelte Transformation, die Inkubation trockener Embryonen in DNA-haltiger Lösung, oder andere Methoden der direkten DNA-Einführung (DE 4 005 152, WO 90/12096, US 4,684,611). Entsprechende Verfahren sind beschrieben (beispielsweise bei Bilang et al. (1991) *Gene* 100:247-250; Scheid et al. (1991) *Mol Gen Genet* 228:104-112; Guerche et al. (1987) *Plant Science* 52:111-116; Neuhaase et al. (1987) *Theor Appl Genet* 75:30-36; Klein et al. (1987) *Nature* 327:70-73; Howell et al. (1980) *Science* 208:1265; Horsch et al. (1985) *Science* 227:1229-1231; DeBlock et al. (1989) *Plant Physiology* 91:694-701; *Methods for Plant Molecular Biology* (Weissbach and Weissbach, eds.) Academic Press Inc. (1988); and *Methods in Plant Molecular Biology* (Schuler and Zielinski, eds.) Academic Press Inc. (1989)). Physikalische Methoden der DNA-Einführung in pflanzliche Zelle sind im Überblick dargestellt bei Oard (1991) *Biotech Adv* 9:1-11.

20

Im Falle dieser "direkten" Transformationsmethoden sind keine besonderen Anforderungen an das verwendete Plasmid gestellt. Einfache Plasmide wie die der pUC-Reihe, pBR322, M13mp Reihe, pACYC184 etc. können verwendet werden.

25

- Neben diesen "direkten" Transformationstechniken kann eine Transformation auch durch bakterielle Infektion mittels *Agrobacterium* (z.B. EP 0 116 718), virale Infektion mittels viraler Vektoren (EP 0 067 553; US 4,407,956; WO 95/34668; WO 93/03161) oder mittels Pollen (EP 0 270 356; WO 85/01856; US 4,684,611) durchgeführt werden.

- Bevorzugt erfolgt die Transformation mittels *Agrobacterien*, die "entwaffnete" (disarmed) Ti-Plasmidvektoren enthalten, wobei deren natürliche Fähigkeit zum Gentransfer auf Pflanzen genutzt wird (EP-A 0 270 355; EP-A 0 116 718). *Agrobacterium*-Transformation ist weit verbreitet für die Transformation von Dicotyledonen, wird aber auch zunehmend auf Monocotyledonen angewandt (Toriyama et al. (1988) *Bio/Technology* 6: 1072-1074; Zhang et al. (1988) *Plant Cell Rep* 7:379-384; Zhang et al. (1988) *Theor Appl Genet* 76:835-840; Shimamoto et al. (1989) *Nature* 338:274-276; Datta et al. (1990) *Bio/Technology* 8: 736-740; Christou et al. (1991) *Bio/Technology* 9:957-962; Peng et al. (1991) International Rice Research Institute, Manila, Philippines 563-574; Cao et al. (1992) *Plant Cell Rep* 11:585-591; Li et al. (1993) *Plant Cell Rep* 12:250-255; Rathore et al. (1993) *Plant Mol Biol* 21:871-884; Fromm et al. (1990) *Bio/Technology* 8:833-839;

- Gordon-Kamm et al. (1990) Plant Cell 2:603-618; D'Halluin et al. (1992) Plant Cell 4:1495-1505; Walters et al. (1992) Plant Mol Biol 18:189-200; Kozziel et al. (1993) Biotechnology 11:194-200; Vasil IK (1994) Plant Mol Biol 25:925-937; Weeks et al. (1993) Plant Physiol 102:1077-1084; Somers et al. (1992) Bio/Technology 10:1589-1594; WO 92/14828; Hiei et al. (1994) Plant J 6:271-282).

Die für die Agrobacterium-Transformation meist verwendeten Stämme Agrobacterium tumefaciens oder Agrobacterium rhizogenes enthalten ein Plasmid (Ti bzw. Ri Plasmid), das auf die Pflanze nach Agrobakterium-Infektion übertragen wird. Ein Teil dieses Plasmids, genannt T-DNA (transferred DNA), wird in das Genom der Pflanzenzelle integriert. Alternativ können durch Agrobakterium auch binäre Vektoren (Mini-Ti-Plasmide) auf Pflanzen übertragen und in deren Genom integriert werden.

Die Anwendung von Agrobacterium tumefaciens für die Transformation von Pflanzen unter Verwendung von Gewebekultur-explantaten ist beschrieben (u.a. Horsch RB et al. (1985) Science 225:1229ff; Fraley et al. (1983) Proc Natl Acad Sci USA 80: 4803-4807; Bevens et al. (1983) Nature 304:184-187). Viele Stämme von Agrobacterium tumefaciens sind in der Lage, genetisches Material zu übertragen, wie z.B. die Stämme EHA101[pEHA101], EHA105[pEHA105], LBA4404[pAL4404], C58C1[pMP90] und C58C1[pGV2260] (Hood et al. (1993) Transgenic Res 2:208-218; Hoekema et al. (1983) Nature 303:179-181; Koncz and Schell (1986) Gen Genet 204:383-396; Deblaere et al. (1985) Nucl Acids Res 13: 4777-4788).

Werden Agrobakterien verwendet, so ist die Expressionskassette in spezielle Plasmide zu integrieren, entweder in einen Zwischenvektor (englisch: shuttle or intermediate vector) oder einen binären Vektor. Wird ein Ti oder Ri Plasmid zur Transformation verwendet, ist zumindest die rechte Begrenzung, meistens jedoch die rechte und die linke Begrenzung der Ti oder Ri Plasmid T-DNA als flankierende Region mit der einzuführenden Expressionskassette verbunden. Bevorzugt werden binäre Vektoren verwendet. Binäre Vektoren können sowohl in E.coli als auch in Agrobacterium replizieren und enthalten die zur Übertragung in ein pflanzliches System erforderlichen Komponenten. Sie enthalten in der Regel ein Selektionsmarkergen für die Selektion transformierter Pflanzen (z.B. das nptII Gen, das eine Resistenz gegen Kanamycin verleiht) und einen Linker oder Polylinker flankiert von der rechten und linken T-DNA Begrenzungssequenz. Außerhalb der T-DNA-Begrenzungssequenz enthalten sie zudem noch einen Selektionsmarker, der eine Selektion transformierter E.coli und/oder Agrobakteria ermöglicht (z.B. das nptIII Gen, das eine Resistenz gegen Kanamycin ver-

leiht). Entsprechende Vektoren können direkt in Agrobakterium transformiert werden (Holsters et al. (1978) Mol Gen Genet 163:181-187).

- 5 Binärvektoren basieren z.B. auf "broad host range"-Plasmiden wie pRK252 (Bevan et al. (1984) Nucl Acid Res 12,8711-8720) und pTJS75 (Watson et al. (1985) EMBO J 4(2):277- 284). Eine grosse Gruppe der verwendeten Binärvektoren leitet sich vom pBIN19 (Bevan et al. (1984) Nucl Acid Res 12:8711-8720) ab. Hajdukiewicz  
 10 et al. entwickelten einen Binärvektor (pPZP), der kleiner und effizienter als die bisher üblichen ist (Hajdukiewicz et al. (1994) Plant Mol Biol 25:989-994). Verbesserte und besondere bevorzugte binäre Vektorsysteme zur Agrobakterium-vermittelten Transformation sind in WO 02/00900 beschrieben.
- 15 Die mit einem solchen Vektor transformierten Agrobakterien können dann in bekannter Weise zur Transformation von Pflanzen, insbesondere von Kulturpflanzen, wie z.B. von Raps, verwendet werden, indem beispielsweise verwundete Blätter oder Blattstücke  
 20 in einer Agrobakterienlösung gebadet und anschliessend in geeigneten Medien kultiviert werden. Die Transformation von Pflanzen durch Agrobakterien ist beschrieben (White FF, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, herausgegeben von S.D. Kung und  
 25 R. Wu, Academic Press, 1993, S. 15-38; Jenes B et al. (1993) Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, herausgegeben von S.D. Kung und R. Wu, Academic Press, S.128-143; Potrykus (1991) Annu Rev Plant Physiol Plant Molec Biol 42:205- 225). Aus den trans-  
 30 formierten Zellen der verwundeten Blätter bzw. Blattstücke können in bekannter Weise transgene Pflanzen regeneriert werden.

Für die Transformation können unterschiedliche Explantate, Zellkulturen, Gewebe, Organen, Embryonen, Samen, Mikrosporen oder  
 35 anderen einzelligen oder mehrzelligen zelluläre Strukturen abgeleitet von einem pflanzlichen Organismus eingesetzt werden. Auf die jeweiligen Explantate, Kulturen oder Gewebe abgestimmte Transformationsverfahren sind dem Fachmann bekannt. Beispielhaft seien zu nennen: Sprossinternodien (Fry J et al. (1987) Plant  
 40 Cell Rep. 6:321-325), Hypokotyle (Radke SE et al. (1988) Theor Appl Genet 75:685-694; Schröder M et al. (1994) Physiologia Plant 92: 37-46.; Stefanov I et al. (1994) Plant Sci. 95:175-186; Weier et al. (1997) Fett/Lipid 99:160-165), kotyledonäre Petiolen (Meloney MM et al. (1989) Plant Cell Rep 8:238-242; Weier D  
 45 et al. (1998) Molecular Breeding 4:39-46), Mikrosporen und Proembryonen (Pechnan (1989) Plant Cell Rep. 8:387-390) und Blütenstiele (Boulter ME et al. (1990) Plant Sci 70:91-99; Guerche P

- et al. (1987) Mol Gen Genet 206:382-386). Bei einem direkten Gentransfer können Mesophyllprotoplasten (Chapel PJ & Glimelius K (1990) Plant Cell Rep 9: 105-108; Golz et al. (1990) Plant Mol Biol 15:475-483) aber auch Hypokotylprotoplasten (Bergmann P & Glimelius K (1993) Physiologia Plant 88:604-611) und Mikrosporen (Chen JL et al. (1994) Theor Appl Genet 88:187-192; Jonesvilleneuve E et al. (1995) Plant Cell Tissue and Organ Cult 40:97-100) und Sprossabschnitte (Seki M et al. (1991) Plant Mol Biol 17:259-263) erfolgreich eingesetzt werden.
- 10 Stabil transformierte Zellen d.h. solche, die die eingeführte DNA integriert in die DNA der Wirtszelle enthalten, können von untransformierten unter Einsatz des erfindungsgemäßen Selektionsverfahrens selektioniert werden. Die erhaltenen Pflanzen können
- 15 in üblicher Weise gezüchtet und gekreuzt werden. Zwei oder mehr Generationen sollten vorzugsweise kultiviert werden, um sicherzustellen, dass die genomische Integration stabil und vererblich ist.
- 20 Sobald eine transformierte Pflanzenzelle hergestellt wurde, kann eine vollständige Pflanze unter Verwendung von dem Fachmann bekannten Verfahren erhalten werden. Hierbei geht man beispielhaft von Kalluskulturen, einzelnen Zellen (z.B. Protoplasten) oder Blattscheiben aus (Vasil et al. (1984) Cell Culture and
- 25 Somatic Cell Genetics of Plants, Vol I, II and III, Laboratory Procedures and Their Applications, Academic Press; Weissbach and Weissbach (1989) Methods for Plant Molecular Biology, Academic Press). Aus diesen noch undifferenzierten Callus-Zellmassen kann die Bildung von Spross und Wurzel in bekannter Weise induziert
- 30 werden. Die erhaltenen Sprösslinge können ausgepflanzt und gezüchtet werden. Entsprechende Verfahren sind beschrieben (Fennell et al. (1992) Plant Cell Rep. 11: 567-570; Stoeger et al. (1995) Plant Cell Rep. 14:273-278; Jahne et al. (1994) Theor Appl Genet 89:525-533).
- 35 Die Wirksamkeit der Expression der transgen exprimierten Nukleinsäuren kann beispielsweise *in vitro* durch Sprossmeristemvermehrung unter Verwendung einer der oben beschriebenen Selektionsmethoden ermittelt werden. Zudem kann eine in Art und Höhe ver-
- 40 änderte Expression eines Zielgens und die Auswirkung auf den Phänotyp der Pflanze an Testpflanzen in Gewächshausversuchen getestet werden.
- Bevorzugt wird das erfindungsgemäße Verfahren im Rahmen der
- 45 Pflanzenbiotechnologie zur Erzeugung von Pflanzen mit vorteilhaften Eigenschaften eingesetzt. Die in das Genom der pflanzlichen Zelle oder des pflanzlichen Organismus zu

"insertierende Nukleinsäuresequenz" umfasst bevorzugt mindestens eine Expressionskassette, wobei besagte Expressionskassette unter Kontrolle eines in pflanzlichen Zellen oder pflanzlichen Organismen funktionellen Promotors eine RNA und/oder ein Protein exprimieren kann, welche nicht die Verminderung der Expression, Menge, Aktivität und/oder Funktion eines Markerproteins bewirken, sondern - besonders bevorzugt - der so genetische veränderten Pflanze einen vorteilhaften Phänotyp verleihen. Dem Fachmann sind zahlreiche Gene und Proteine bekannt, die zum Erreichen eines vorteilhaften Phänotyp beispielsweise zur Qualitätssteigerung bei Nahrungsmitteln oder zur Produktion bestimmter Chemikalien oder Pharmazeutika (Dunwell JM (2000) J Exp Bot 51 Spec No:487-96) verwendet werden können.

So kann die Eignung der Pflanzen oder deren Samen als Nahrungs- oder Futtermittel verbessert werden, beispielsweise über eine Veränderung der Zusammensetzungen und/oder des Gehalt an Metaboliten, insbesondere Proteinen, Ölen, Vitaminen und/oder Stärke. Auch können Wachstumsrate, Ertrag oder die Resistenz gegen biotische oder abiotische Stressfaktoren erhöht werden. Vorteilhafte Effekte können sowohl durch transgene Expression von Nukleinsäuren oder Proteinen als auch durch gezielte Verminderung der Expression endogener Gene hinsichtlich des Phänotypes der transgenen Pflanze erzielt werden. Die in der transgenen Pflanze zu erzielenden vorteilhaften Effekte umfassen beispielsweise:

- Erhöhte Resistenz gegen Pathogene (biotischer Stress)
- Erhöhte Resistenz gegen Umwelteinflüsse wie Hitze, Kälte, Frost Trockenheit, UV-Licht, oxidativen Stress, Nässe, Salz etc. (abiotischer Stress)
- Erhöhte Ertragsleistung
- Verbesserte Qualität z.B. erhöhter Nährwert, erhöhte Lagerfähigkeit

Ein weiterer Gegenstand der Erfindung betrifft die Verwendung der nach dem erfindungsgemässen Verfahren hergestellten transgenen Pflanzen, und der von ihnen abgeleitete Zellen, Zellkulturen, Teile oder Vermehrungsgut wie Saaten oder Früchte, zur Herstellung von Nahrungs- oder Futtermitteln, Pharmazeutika oder Feinchemikalien, wie beispielsweise Enzymen, Vitaminen, Aminosäuren, Zucker, Fettsäuren, natürliche und synthetische Geschmacks-, Aroma- und Farbstoffe. Besonders bevorzugt ist die Produktion von Triacylglyceriden, Lipiden, Ölen, Fettsäuren, Stärke, Tocopherolen und Tocotrienolen sowie Carotinoiden. Von

Menschen und Tieren verzehrbare erfindungsgemäße, genetisch veränderte Pflanzen können auch beispielsweise direkt oder nach an sich bekannter Aufbereitung als Nahrungsmittel oder Futtermittel verwendet werden.

- 5 Wie bereits oben erwähnt, umfasst das erfindungsgemäße Verfahren in einer besonders vorteilhaften Ausführungsform in einem der Selektion nachgeschalteten Verfahrensschritt die Deletion der für das Markerprotein kodierenden Sequenz (z.B. durch Rekombinase vermittelt oder wie in WO03/004659 beschrieben) bzw. die Aus-
- 10 kreuzung und/oder Segregation besagter Sequenzen. (Dem Fachmann ist klar, das dazu in den transformierten Zellen die in das Genom integrierte Nukleinsäuresequenz und die für das Markerprotein kodierende Sequenz einen separaten chromosomalen Lokus aufweisen sollten. Dies ist jedoch bei der Mehrzahl der resultierenden
- 15 Pflanzen allein aus statistischen Gründen gegeben). Diese Vorgehensweise ist insbesondere vorteilhaft, wenn es sich bei dem Markerprotein um ein Transgen handelt, das ansonsten in der zu transformierenden Pflanze nicht vorkommt. Die resultierende Pflanze kann zwar noch unter Umständen die Verbindung zur Ver-
- 20 minderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins beinhalten, doch hätte diese kein "Gegenstück" mehr in Form des besagten Markerproteins, wäre also wirkungslos. Dies ist insbesondere dann der Fall, wenn das Markerprotein aus einem nicht-pflanzlichen Organismus stammt und/oder synthetisch
- 25 ist (beispielsweise das *codA* Protein). Es können aber auch Pflanzliche Markerproteine aus anderen Pflanzenarten eingesetzt werden, die in der zu transformierenden Zelle ansonsten (d.h. wenn nicht als Transgen eingeführt) nicht vorkommen. Besagte Markerproteine sind im Rahmen dieser Erfindung als "nicht-endo-
- 30 gene" Markerproteine bezeichnet.

- Ganz besonders vorteilhaft ist es, wenn es sich bei der Verbindung zur Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins um eine RNA handelt. Nach
- 35 der Deletion bzw. Auskreuzung/Segregation würde die resultierende transgene Pflanze kein unnötiges (und ggf. unerwünschtes) Fremdprotein mehr aufweisen. Einziges Fremdprotein ist u.U. das aus der in das Genom insertierten Nukleinsäuresequenz resultierenden Protein. Aus Gründen der Produktzulassung ist diese Ausführungs-
- 40 form besonders vorteilhaft. Wie oben beschrieben kann diese RNA eine antisense RNA oder - besonders bevorzugt - eine doppelsträngige RNA sein. Sie kann separat von der für das Zielprotein kodierenden RNA exprimiert werden, aber - unter Umständen - auch auf einem Strang mit derselben.

## 66

Zusammenfassend umfasst die besonders vorteilhafte Ausführungsform folgende Merkmale:

Verfahren zur Herstellung transformierter pflanzlicher Zellen  
5 oder Organismen, umfassend nachfolgende Schritte:

- 10 a) Transformation einer Population pflanzlicher Zellen, welche mindestens ein nicht-endogenes (bevorzugt nicht-pflanzliches) Markerprotein umfasst, das in der Lage ist, eine für besagte Population pflanzlicher Zellen nicht-toxische Substanz X in  
15 eine für besagte Population toxische Substanz Y direkt oder indirekt umzusetzen, mit mindestens einer zu insertierenden Nukleinsäuresequenz in Kombination mit mindestens einer Nukleinsäuresequenz kodierend für eine Ribonukleinsäuresequenz befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion besagten Markerproteins, und
- 20 b) Behandlung besagter Population pflanzlicher Zellen mit der Substanz X in einer Konzentration, die infolge der Umsetzung durch das Markerprotein einen für nicht-transformierte Zellen toxischen Effekt bedingt, und
- 25 c) Selektion von transformierten pflanzlichen Zellen (und/oder Populationen pflanzlicher Zellen wie pflanzlichen Geweben oder Pflanzen), die in ihrem Genom besagte Nukleinsäuresequenz aufweisen und die infolge der Wirkung besagter Verbindung gegenüber nicht-transformierten Zellen einen Wachstumsvorteil haben, aus besagter Population pflanzlicher Zellen, wobei die Selektion unter Bedingungen durchgeführt  
30 wird, bei denen das Markerprotein seinen toxischen Effekt auf die nicht-transformierten Zellen ausüben kann, und
- 35 d) Regeneration von fertilen Pflanzen, und
- e) Auskreuzung der für das Markerprotein kodierenden Nukleinsäuresequenz und Isolation von fertilen Pflanzen, die in ihrem Genom besagte Nukleinsäuresequenz aber nicht mehr die für das Markerprotein kodierende Sequenz aufweisen.

## 40 Sequenzen

1. SEQ ID NO: 1 Nukleinsäuresequenz kodierend für Cytosindeaminase aus E.coli (codA)
- 45 2. SEQ ID NO: 2 Aminosäuresequenz kodierend für Cytosindeaminase aus E.coli (codA)

## 67

3. SEQ ID NO: 3 Nukleinsäuresequenz kodierend für Cytosin-deaminase aus E.coli (codA) mit modifiziertem Startkodon (GTG/ATG) zur Expression in Eukaryoten
- 5
4. SEQ ID NO: 4 Aminosäuresequenz kodierend für Cytosin-deaminase aus E.coli (codA) mit modifiziertem Startkodon (GTG/ATG) zur Expression in Eukaryoten
- 10
5. SEQ ID NO: 5 Nukleinsäuresequenz kodierend für Cytochrom P450-SU1 (suaC) aus Streptomyces griseolus
- 15
6. SEQ ID NO: 6 Aminosäuresequenz kodierend für Cytochrom P450-SU1 (suaC) aus Streptomyces griseolus
7. SEQ ID NO: 7 Nukleinsäuresequenz kodierend für Indolacetamidhydrolase (tms2) aus Agrobacterium tumefaciens
- 20
8. SEQ ID NO: 8 Aminosäuresequenz kodierend für Indolacetamidhydrolase (tms2) aus Agrobacterium tumefaciens
- 25
9. SEQ ID NO: 9 Nukleinsäuresequenz kodierend für Indolacetamidhydrolase (tms2) aus Agrobacterium tumefaciens
- 30
10. SEQ ID NO: 10 Aminosäuresequenz kodierend für Indolacetamidhydrolase (tms2) aus Agrobacterium tumefaciens
- 35
11. SEQ ID NO: 11 Nukleinsäuresequenz kodierend für Haloalkan-dehalogenase (dhlA) aus Xanthobacter autotrophicus
- 40
12. SEQ ID NO: 12 Aminosäuresequenz kodierend für Haloalkan-dehalogenase (dhlA) aus Xanthobacter autotrophicus
- 45
13. SEQ ID NO: 13 Nukleinsäuresequenz kodierend für Thymidin-kinase aus Herpes simplex Virus 1
14. SEQ ID NO: 14 Aminosäuresequenz kodierend für Thymidin-kinase aus Herpes simplex Virus 1



## 68

15. SEQ ID NO: 15 Nukleinsäuresequenz kodierend für Thymidin-kinase aus Herpes simplex Virus 1
- 5 16. SEQ ID NO: 16 Aminosäuresequenz kodierend für Thymidin-kinase aus Herpes simplex Virus 1
- 10 17. SEQ ID NO: 17 Nukleinsäuresequenz kodierend für Hypoxanthin-Xanthin-Guanin Phosphoribosyltransferase aus *Toxoplasma gondii*
18. SEQ ID NO: 18 Aminosäuresequenz kodierend für Hypoxanthin-Xanthin-Guanin Phosphoribosyltransferase aus *Toxoplasma gondii*
- 15 19. SEQ ID NO: 19 Nukleinsäuresequenz kodierend für Xanthin-Guanin-Phosphoribosyltransferase aus *E.coli*
20. SEQ ID NO: 20 Aminosäuresequenz kodierend für Xanthin-Guanin-Phosphoribosyltransferase aus *E.coli*
- 20 21. SEQ ID NO: 21 Nukleinsäuresequenz kodierend für Xanthin-Guanin-Phosphoribosyltransferase aus *E.coli*
22. SEQ ID NO: 22 Aminosäuresequenz kodierend für Xanthin-Guanin-Phosphoribosyltransferase aus *E.coli*
- 25 23. SEQ ID NO: 23 Nukleinsäuresequenz kodierend für Purin-nukleosidphosphorylase (deoD) aus *E.coli*
- 30 24. SEQ ID NO: 24 Nukleinsäuresequenz kodierend für Purin-nukleosidphosphorylase (deoD) aus *E.coli*
25. SEQ ID NO: 25 Nukleinsäuresequenz kodierend für Phosphonat-monoesterhydrolase (pehA) aus *Burkholderia caryophylli*
- 35 26. SEQ ID NO: 26 Aminosäuresequenz kodierend für Phosphonat-monoesterhydrolase (pehA) aus *Burkholderia caryophylli*
- 40 27. SEQ ID NO: 27 Nukleinsäuresequenz kodierend für Tryptophan-oxygenase (aux1) aus *Agrobacterium rhizogenes*
- 45 28. SEQ ID NO: 28 Aminosäuresequenz kodierend für Tryptophan-oxygenase (aux1) aus *Agrobacterium rhizogenes*

## 69

29. SEQ ID NO: 29 Nukleinsäuresequenz kodierend für Indol-acetamidhydrolase (aux2) aus *Agrobacterium rhizogenes*
- 5 30. SEQ ID NO: 30 Aminosäuresequenz kodierend für Indol-acetamidhydrolase (aux2) aus *Agrobacterium rhizogenes*
- 10 31. SEQ ID NO: 31 Nukleinsäuresequenz kodierend für Tryptophan-oxygenase (aux1) aus *Agrobacterium tumefaciens*
32. SEQ ID NO: 32 Aminosäuresequenz kodierend für Tryptophan-oxygenase (aux1) aus *Agrobacterium tumefaciens*
- 15 33. SEQ ID NO: 33 Nukleinsäuresequenz kodierend für Indol-acetamidhydrolase (aux2) aus *Agrobacterium tumefaciens*
- 20 34. SEQ ID NO: 34 Aminosäuresequenz kodierend für Indol-acetamidhydrolase (aux2) aus *Agrobacterium tumefaciens*
- 25 35. SEQ ID NO: 35 Nukleinsäuresequenz kodierend für Indol-acetamidhydrolase (aux2) aus *Agrobacterium vitis*
- 30 36. SEQ ID NO: 36 Aminosäuresequenz kodierend für Indol-acetamidhydrolase (aux2) aus *Agrobacterium vitis*
37. SEQ ID NO: 37 Nukleinsäuresequenz kodierend für 5-Methylthioribosekinase (mtrK) aus *Arabidopsis thaliana*
- 35 38. SEQ ID NO: 38 Aminosäuresequenz kodierend für 5-Methylthioribosekinase (mtrK) aus *Arabidopsis thaliana*
- 40 39. SEQ ID NO: 39 Nukleinsäuresequenz kodierend für 5-Methylthioribosekinase (mtrK) aus *Klebsiella pneumoniae*
- 45 40. SEQ ID NO: 40 Aminosäuresequenz kodierend für 5-Methylthioribosekinase (mtrK) aus *Klebsiella pneumoniae*

## 70

41. SEQ ID NO: 41 Nukleinsäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Arabidopsis thaliana*
- 5 42. SEQ ID NO: 42 Aminosäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Arabidopsis thaliana*
43. SEQ ID NO: 43 Nukleinsäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Hordeum vulgare*  
(Gerste)
- 10 44. SEQ ID NO: 44 Aminosäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Hordeum vulgare*  
(Gerste)
- 15 45. SEQ ID NO: 45 Nukleinsäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Oryza sativa* (Reis)
46. SEQ ID NO: 46 Aminosäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Oryza sativa* (Reis)
- 20 47. SEQ ID NO: 47 Nukleinsäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Zea mays* (Mais)
48. SEQ ID NO: 48 Aminosäuresequenz kodierend für Alkohol-  
dehydrogenase (adh) aus *Zea mays* (Mais)
- 25 49. SEQ ID NO: 49 Nukleinsäuresequenz kodierend für ein sense-RNA  
Fragment der Cytosindeaminase aus *E.coli*  
(codARNai-sense)
- 30 50. SEQ ID NO: 50 Oligonukleotidprimer codA5'HindIII  
5'-AAGCTTGGCTAACAGTGTCGAATAACG-3'
51. SEQ ID NO: 51 Oligonukleotidprimer codA3'SalI  
5'-GTGACGACAAAATCCCTTCCTGAGG-3'
- 35 52. SEQ ID NO: 52 Nukleinsäuresequenz kodierend für ein  
antisense-RNA Fragment der Cytosindeaminase aus  
*E.coli* (codARNai-anti)
- 40 53. SEQ ID NO: 53 Oligonukleotidprimer codA5'EcoRI  
5'-GAATTCGGCTAACAGTGTCGAATAACG-3'
54. SEQ ID NO: 54 Oligonukleotidprimer codA3'BamHI  
5'-GGATCCGACAAAATCCCTTCCTGAGG-3'
- 45

## 71

55. SEQ ID NO: 55 Vektorkonstrukt pBluKS-nitP-STLS1-35S-T
56. SEQ ID NO: 56 Expressionsvektor pSUN-1
- 5 57. SEQ ID NO: 57 Transgener Expressionsvektor pSUN-1-codA-RNAi
58. SEQ ID NO: 58 Transgener Expressionsvektor  
pSUN1-codA-RNAi-At.Act.-2-At.Als-R-ocsT
- 10 59. SEQ ID NO: 59 Nukleinsäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Mais (Zea mays);  
Fragment
60. SEQ ID NO: 60 Aminosäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Mais (Zea mays);  
15 Fragment
61. SEQ ID NO: 61 Nukleinsäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Raps (Brassica  
napus), Fragment
- 20 62. SEQ ID NO: 62 Aminosäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Raps (Brassica  
napus), Fragment
- 25 63. SEQ ID NO: 63 Nukleinsäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Raps (Brassica  
napus), Fragment
64. SEQ ID NO: 64 Aminosäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Raps (Brassica  
30 napus), Fragment
65. SEQ ID NO: 65 Nukleinsäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Reis (Oryza sativa)  
35 Fragment
66. SEQ ID NO: 66 Aminosäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Reis (Oryza sativa)  
Fragment
- 40 67. SEQ ID NO: 67 Nukleinsäuresequenz kodierend für 5-Methyl-  
thioribosekinase (mtrK) aus Soja (Glycine max),  
Fragment
- 45

## 72

68. SEQ ID NO: 68 Aminosäuresequenz kodierend für 5-Methylthioribosekinase (mtrK) aus Soja (Glycine max),  
Fragment

5 69. SEQ ID NO: 69 Oligonukleotidprimer codA5'C-term  
5'-CGTGAATACGGCGTGGAGTCG-3'

70. SEQ ID NO: 70 Oligonukleotidprimer codA3'C-term  
5'-CGGCAGGATAATCAGGTTGG-3'

10

71. SEQ ID NO: 71 Oligonukleotidprimer 35sT 5' Primer  
5'-GTCAACGTAACCAACCCTGC-3'

15

20

25

30

35

40

45

## Abbildungen

Fig.1: Inaktivierung des Markerproteingens mittels Einbringen einer Rekombinase

5

P: Promotor  
 MP: Sequenz kodierend für ein Markerprotein  
 R1/R2: Rekombinase-Erkennungssequenzen  
 R: Rekombinase bzw. Sequenz kodierend für  
 10 Rekombinase.

15

In einer bevorzugten Ausführungsform wird die Inaktivierung des Markerproteingens durch das Einbringen einer sequenzspezifischen Rekombinase realisiert. Bevorzugt wird die Rekombinase - wie hier dargestellt - ausgehend von einer Expressionskassette exprimiert.

20

Das Markerproteingen ist von Erkennungssequenzen für sequenzspezifische Rekombinasen flankiert, wobei durch Einbringen der Rekombinase Sequenzen des Markerproteingens deletiert werden und so eine Inaktivierung des Markerproteingens erfolgt.

25

Fig.2-A: Inaktivierung des Markerproteingens durch Einwirken einer sequenzspezifischen Nuklease

30

P: Promotor  
 DS: Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen  
 MP-DS-MP': Sequenz kodierend für ein Markerprotein umfassend eine DS  
 nDS: Inaktivierte DS  
 E: Sequenzspezifisches Enzym zur gezielten Induktion von DNA-Doppelstrangbrüchen

35

40

45

Das Markerprotein-Gen kann durch eine gezielte Mutation oder Deletion im Markerprotein-Gen z.B. durch sequenzspezifische Induktion von DNA-Doppelstrangbrüchen an einer Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen in oder in der Nähe des Markerprotein-Gens (P-MP) realisiert werden. Der Doppelstrangbruch kann in der kodierenden Region oder aber auch der nicht-kodierenden (wie beispielsweise dem Promotor) erfolgen, induziert eine illegitime Rekombination (nicht-homologe Verbindung von DNA-Enden; "non-homologous end-

joining") und so z.B. eine Verschiebung im Leseraster des Markerproteins.

Fig.2-B: Inaktivierung des Markerproteingens durch Einwirken einer sequenzspezifischen Nuklease

5

P: Promotor  
 DS: Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen  
 10 MP: Sequenz kodierend für ein Markerprotein  
 nDS: Inaktivierte DS  
 E: Sequenzspezifisches Enzym zur gezielten Induktion von DNA-Doppelstrangbrüchen

15 Das Markerprotein-Gen kann durch eine gezielte Deletion durch sequenzspezifische Induktion von mehr als einem sequenzspezifischen DNA-Doppelstrangbruch in oder in der Nähe des Markerprotein-Gens realisiert werden. Die Doppelstrangbrüche können in der kodierenden Region  
 20 oder aber auch der nicht-kodierenden (wie beispielsweise dem Promotor) erfolgen und induzieren eine Deletion im Markerprotein-Gen. Bevorzugt ist das Markerprotein-Gen von DS Sequenzen flankiert und wird vollständig durch Einwirken des Enzyms E deletiert.

25

Fig. 3: Inaktivierung des Markerprotein-Gens durch Induktion einer intramolekularen homologen Rekombination infolge des Einwirkens einer sequenzspezifischen Nuklease

30 A/A': Sequenzen mit einer ausreichenden Länge und Homologie zueinander, um infolge des induzierten Doppelstrangbruches miteinander zu rekombinieren

P: Promotor  
 35 DS: Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen  
 MP: Sequenz kodierend für ein Markerprotein  
 E: Sequenzspezifisches Enzym zur gezielten Induktion von DNA-Doppelstrangbrüchen

40 Das Markerprotein-Gen kann durch eine Deletion mittels intramolekularen homologer Rekombination inaktiviert werden. Die homologe Rekombination kann durch sequenzspezifische Induktion von DNA-Doppelstrangbrüchen an  
 45 einer Erkennungssequenz zur gezielten Induktion von DNA- Doppelstrangbrüchen in oder in der Nähe des Markerprotein-Gens initiiert werden. Die homologe Rekombination

## 75

erfolgt zwischen den Sequenzen A und A', die eine ausreichenden Länge und Homologie zueinander haben, um infolge des induzierten Doppelstrangbruches miteinander zu rekombinieren. Die Rekombination bewirkt eine Deletion essentieller sequenzen des Markerprotein-Gens.

5

Fig. 4: Inaktivierung des Markerprotein-Gens durch intermolekulare homologe Rekombination

10 A/A': Sequenzen mit einer ausreichenden Länge und Holomogie zueinander, um miteinander zu rekombinieren  
 B/B': Sequenzen mit einer ausreichenden Länge und Holomogie zueinander, um miteinander zu rekombinieren

15 P: Promotor  
 I: zu insertierende Nukleinsäuresequenz / Gen von Interesse  
 MP: Sequenz kodierend für ein Markerprotein

20 Die Inaktivierung des Markerprotein-Gens (P-MP) kann auch durch eine gezielte Insertion in das Markerprotein-Gen z.B. mittels intermolekularer homologer Rekombination realisiert werden. Dabei ist die zu insertierende Region an ihrem 5'- und 3'-Ende von Nukleinsäuresequenzen (A' bzw. B') flankiert, die eine ausreichende Länge und Homologie zu entsprechenden flankierenden Sequenzen des Markerproteingens (A bzw. B) aufweisen, um eine homologe Rekombination zwischen A und A' und B und B' zu ermöglichen. Die Rekombination bewirkt eine Deletion essentieller sequenzen des Markerprotein-Gens.

25  
30

Fig. 5: Inaktivierung des Markerprotein-Gens durch intermolekulare homologe Rekombination infolge des Einwirkens einer sequenzspezifischen Nuklease

35

A/A': Sequenzen mit einer ausreichenden Länge und Holomogie zueinander, um miteinander zu rekombinieren

40 B/B': Sequenzen mit einer ausreichenden Länge und Holomogie zueinander, um miteinander zu rekombinieren

P: Promotor  
 I: zu insertierende Nukleinsäuresequenz / Gen von Interesse

45 MP: Sequenz kodierend für ein Markerprotein  
 DS: Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen



## 76

E: Sequenzspezifisches Enzym zur gezielten  
Induktion von DNA-Doppelstrangbrüchen

Die Inaktivierung des Markerprotein-Gens kann auch durch  
eine gezielte Insertion in das Markerprotein-Gen z.B.  
mittels intermolekularer homologer Rekombination reali-  
siert werden. Die homologe Rekombination kann durch  
sequenzspezifische Induktion von DNA-Doppelstrangbrüchen  
an einer Erkennungssequenz zur gezielten Induktion von  
DNA-Doppelstrangbrüchen in oder in der Nähe des Marker-  
protein-Gens initiiert werden. Dabei ist die zu inser-  
tierende Region an ihrem 5'- und 3'-Ende von Nuklein-  
säuresequenzen (A' bzw. B') flankiert, die eine aus-  
reichende Länge und Homologie zu entsprechenden flan-  
kierenden Sequenzen des Markerprotein-Gens (A bzw. B)  
aufweisen, um eine homologe Rekombination zwischen A und  
A' und B und B' zu ermöglichen. Die Rekombination bewirkt  
eine Deletion essentieller sequenzen des Markerprotein-  
Gens.

Fig. 6: Vektorkarte für pBluKS-nitP-STLS1-35S-T (SEQ ID NO: 55)

NitP: Promotor des NitrilaseI-Gens aus *A.thaliana* (Gen-  
Bank Acc.-No.: Y07648.2, Hillebrand et al. (1996) Gene  
170:197-200)

STLS-1 Intron: Intron des ST-LS1 Gens aus Kartoffel (Van-  
canneyt GF et al. (1990) Mol Gen Genet 220(2):245-250).

35S-Term: Terminator des 35S CaMV Gens (Blumenkohlmosaik-  
virus; Franck et al. (1980) Cell 21:285-294).

Schnittstellen relevanter Restriktionsendonukleasen sind  
mit ihrer jeweiligen Schnittposition angegeben.

Fig. 7: Vektorkarte für den transgenen Expressionsvektor  
pSUN-1-codA-RNAi (SEQ ID NO: 57)

NitP: Promotor des NitrilaseI-Gens aus *A.thaliana* (Gen-  
Bank Acc.-No.: Y07648.2, Hillebrand et al. (1996) Gene  
170:197-200)

STLS-1 Intron: Intron des ST-LS1 Gens aus Kartoffel (Van-  
canneyt GF et al. (1990) Mol Gen Genet 220(2):245-250).

## 77

35S-Term: Terminator des 35S CaMV Gens (Blumenkohlmosaikvirus; Franck et al. (1980) Cell 21:285-294).

5       codA-sense: Nukleinsäuresequenz kodierend für ein sense-RNA Fragment der Cytosindeaminase aus E.coli (codARNAi-sense; SEQ ID NO: 49)

10       codA-anti: Nukleinsäuresequenz kodierend für ein anti-sense-RNA Fragment der Cytosindeaminase aus E.coli (codARNAi-anti; SEQ ID NO: 52)

LB/RB: Linke bzw. rechte Grenze der Agrobacterium T-DNA

15       Schnittstellen relevanter Restriktionsendonukleasen sind mit ihrer jeweiligen Schnittposition angegeben. Weitere Elemente stellen übliche Elemente eines binären Agrobakterium-Vektors dar (aadA; Cole1; repA)

20       Fig. 8: Vektorkarte für den transgenen Expressionsvektor pSUN1-codA-RNAi-At.Act.-2-At.Als-R-ocsT (SEQ ID NO: 58)

25       NitP: Promotor des Nitrilase I-Gens aus A.thaliana (Gen-Bank Acc.-No.: Y07648.2, Hillebrand et al. (1996) Gene 170:197-200)

STLS-1 Intron: Intron des ST-LS1 Gens aus Kartoffel (Vancanneyt GF et al. (1990) Mol Gen Genet 220(2):245-250).

30       35S-Terminator: Terminator des 35S CaMV Gens (Blumenkohlmosaikvirus; Franck et al. (1980) Cell 21:285-294).

35       codA-sense: Nukleinsäuresequenz kodierend für ein sense-RNA Fragment der Cytosindeaminase aus E.coli (codARNAi-sense; SEQ ID NO: 49)

codA-anti: Nukleinsäuresequenz kodierend für ein anti-sense-RNA Fragment der Cytosindeaminase aus E.coli (codARNAi-anti; SEQ ID NO: 52)

40       Left Border/Right Border: Linke bzw. rechte Grenze der Agrobacterium T-DNA

45       Schnittstellen relevanter Restriktionsendonukleasen sind mit ihrer jeweiligen Schnittposition angegeben. Weitere Elemente stellen übliche Elemente eines binären Agrobakterium-Vektors dar (aadA; Cole1; repA)

## 78

Fig.9a-b: Sequenzvergleich von diversen 5-Methylthioribose (MTR)  
kinasen aus verschiedenen Organismen, insbesondere  
pflanzlichen Organismen. Gezeigt sind Sequenzen aus  
Klebsiella pneumoniae, Clostridium tetani, Arabidopsis  
thaliana (A.thaliana), Raps (Brassica napus), Sojabohne  
(Soy-1), Reis (Oryza sativa-1) sowie die Konsensus-  
sequenz (Consensus). Homologe Bereiche können aus der  
Konsensussequenz leicht abgeleitet werden.

5

10

15

20

25

30

35

40

45

## Ausführungsbeispiele

## Allgemeine Methoden

- 5 Die chemische Synthese von Oligonukleotiden kann beispielsweise, in bekannter Weise, nach der Phosphoamiditmethode (Voet, Voet, 2. Auflage, Wiley Press New York, Seite 896-897) erfolgen. Die im Rahmen der vorliegenden Erfindung durchgeführten Klonierungsschritte wie z.B. Restriktionsspaltungen, Agarosegelelektrophorese, Reinigung von DNA-Fragmenten, Transfer von Nukleinsäuren auf Nitrozellulose und Nylonmembranen, Verknüpfen von DNA-Fragmenten, Transformation von *E. coli* Zellen, Anzucht von Bakterien, Vermehrung von Phagen und Sequenzanalyse rekombinanter DNA werden wie bei Sambrook et al. (1989) Cold Spring Harbor Laboratory Press; ISBN 0-87969-309-6 beschrieben durchgeführt. Die Sequenzierung rekombinanter DNA-Moleküle erfolgt mit einem Laserfluoreszenz-DNA-Sequenzierer der Firma ABI nach der Methode von Sanger (Sanger et al. (1977) Proc Natl Acad Sci USA 74:5463-5467).

20

Beispiel 1: Herstellung der *codA*-Fragmente

- Zunächst wird eine verkürzte und am 5' bzw. 3' Ende durch Addition von Erkennungssequenzen der Restriktionsenzyme HindIII und SalI modifizierte Nukleinsäurevariante des *codA* Gens unter Verwendung der PCR-Technologie hergestellt. Dazu wird ein Teil des *codA* Gens (GeneBank Acc.-No.: S56903; SEQ ID NO: 1) mittels Polymerasekettenreaktion (PCR) aus dem Herkunftsorganismus *E. coli* unter Verwendung eines sense-spezifischen Primers (*codA*5'HindIII; SEQ ID NO: 50) und eines antisense-spezifischen Primers (*codA*3'SalI; SEQ ID NO: 51) amplifiziert.

*codA*5'HindIII: 5'-AAGCTTGGCTAACAGTGTCGAATAACG-3' (SEQ ID NO: 50)

- 35 *codA*3'SalI: 5'-GTCGACGACAAAATCCCTTCCTGAGG-3' (SEQ ID NO: 51)

Die PCR erfolgt in einem 50 µl Reaktionsansatz in dem enthalten sind:

- 2 µl (200 ng) genomische DNA von *E. coli*
- 40 - 0,2 mM dATP, dTTP, dGTP, dCTP
- 1,5 mM Mg(OAc)<sub>2</sub>
- 5 µg Rinderserum-Albumin
- 40 pmol Primer "*codA*5'HindIII"
- 40 pmol Primer *codA*3'SalI
- 45 - 15 µl 3,3x rTth DNA Polymerase XLPuffer (PE Applied Biosystems)
- 5U rTth DNA Polymerase XL (PE Applied Biosystems)

## 80

Die PCR wird unter folgenden Zyklus-Bedingungen durchgeführt:

- Schritt 1: 5 Minuten 94°C (Denaturierung)  
Schritt 2: 3 Sekunden 94°C  
5 Schritt 3: 1 Minute 60°C (Annealing)  
Schritt 4: 2 Minuten 72°C (Elongation)
- 30 Wiederholungen der Schritte 2 bis 4
- 10 Schritt 5: 10 Minuten 72°C (Post-Elongation)  
Schritt 6: 4°C (Warteschleife)

- Das Amplifikat (codARNai-sense; SEQ ID NO: 49) wird unter Verwendung von Standardmethoden in den PCR Klonierungsvektor pGEM-T  
15 (Promega) kloniert. Die Identität des erzeugten Amplikons wird durch Sequenzierung unter Verwendung des M13F (-40) Primers bestätigt.

- Ein weiteres verkürztes und am 5'- bzw. 3'-Ende durch Addition  
20 von Erkennungssequenzen der Restriktionsenzyme EcoRI und BamHI modifiziertes Fragment des *codA* Gens wird mittels Polymerasekettenreaktion (PCR) aus *E.coli* unter Verwendung eines sense-spezifischen Primers (codA5'EcoRI; SEQ ID NO: 53) und eines antisense-spezifischen Primers (codA3'BamHI; SEQ ID NO: 54)  
25 amplifiziert.

codA5'EcoRI: 5'-GAATTCGGCTAACAGTGTCTGAATAACG-3' (SEQ ID NO: 53)

codA3'BamHI: 5'-GGATCCGACAAAATCCCTTCCTGAGG-3' (SEQ ID NO: 54)

30

Die PCR erfolgt in einem 50 µl Reaktionsansatz in dem enthalten sind:

- 2 µl (200 ng) genomische DNA von *E.coli*
- 0,2 mM dATP, dTTP, dGTP, dCTP
- 35 - 1,5 mM Mg(OAc)<sub>2</sub>
- 5 µg Rinderserum-Albumin
- 40 pmol Primer "codA5'EcoRI"
- 40 pmol Primer "codA3'BamHI"
- 15 µl 3,3x rTth DNA Polymerase XLPuffer (PE Applied Biosystems)
- 40 - 5U rTth DNA Polymerase XL (PE Applied Biosystems)

45

## 81

Die PCR wird unter folgenden Zyklus Bedingungen durchgeführt:

- Schritt 1: 5 Minuten 94°C (Denaturierung)
- Schritt 2: 3 Sekunden 94°C
- 5 Schritt 3: 1 Minute 60°C (Annealing)
- Schritt 4: 2 Minuten 72°C (Elongation)
- 30 Wiederholungen der Schritte 2 bis 4
- 10 Schritt 5: 10 Minuten 72°C (Post-Elongation)
- Schritt 6: 4°C (Warteschleife)

Das Amplifikat (codARNai-anti; SEQ ID NO: 52) wird unter Verwendung von Standardmethoden in den PCR Klonierungsvektor pGEM-T (Promega) kloniert. Die Identität des erzeugten Amplikons wird durch Sequenzierung unter Verwendung des M13F (-40) Primers bestätigt.

Beispiel 2 Herstellung des transgenen Expressionsvektors zur Expression einer codA doppelsträngigen RNA

Die in Beispiel 1 generierten codA-Fragmente werden zur Herstellung eines DNA Konstruktes geeignet zur Expression einer doppelsträngigen codA-RNA verwendet (pSUN-codA-RNAi). Das Konstrukt ist geeignet zur Reduktion der RNA Fließgleichgewichtsmenge (RNA-stady state level) des codA Gens in transgenen Pflanzen und einer daraus resultierenden Unterdrückung der Expression des codA Gens unter Verwendung der „doublestrand RNA interference“ (dsRNAi) Technologie. Die codA RNAi Kassette wird dazu zunächst in dem Plasmid pBluKS-nitP-STLS1-35S-T aufgebaut und anschließend in einem weiteren Klonierungsschritt vollständig in das pSUN-1 Plasmid überführt.

Der Vektor pBluKS-nitP-STLS1-35S-T (SEQ ID NO: 55) ist ein Derivat des pBluescript KS (Stratagene) und enthält den Promotor des NitrilaseI-Gens aus *A.thaliana* (GenBank Acc.-No.: Y07648.2, Nukleotide 2456 bis 4340, Hillebrand et al. (1996) *Gene* 170:197-200), das STLS-1 Intron (Vancanneyt GF et al. (1990) *Mol Gen Genet* 220(2):245-250), Restriktionsschnittstellen die das Intron an der 5'- bzw. 3'-Seite flankieren und eine gerichtete Insertion von DNA Fragmenten ermöglichen, sowie den Terminator des 35S CaMV Gens (Blumenkohlmosaikvirus; Franck et al. (1980) *Cell* 21:285-294). Unter Verwendung dieser Restriktionsschnittstellen (HindIII, SalI, EcoRI, BamHI) werden die Fragmente codARNAi-sense (SEQ ID NO: 49) und codARNai-anti (SEQ ID NO: 52) in

## 82

diesen Vektor inseriert, wodurch die fertige *codA* RNAi Kasse-  
ette entsteht.

Zu diesem Zweck wird zunächst das *codA*-sense Fragment (*codA*RNAi-  
5 sense SEQ ID NO: 49) unter Verwendung der Enzyme HindIII und SalI  
aus dem pGEM-T Vektor herausgeschnitten, isoliert und in den  
pBluKS-nitP-STLS1-35S-T Vektor unter Standardbedingungen ligiert.  
Dieser Vektor wurde im Vorfeld unter Verwendung der Restriktions-  
enzyme HindIII und SalI geschnitten. Entsprechend positive Klone  
10 werden durch analytischen Restriktionsverdau und Sequenzierung  
identifiziert.

Der entstandene Vektor (pBluKS-nitP-*codA*sense-STLS1-35S-T) wird  
unter Verwendung der Restriktionsenzyme BamHI und EcoRI verdaut.  
15 Das *codA*-anti Fragment (*codA*RNAi-anti; SEQ ID NO: 52) wird aus  
dem entsprechenden pGEM-T Vektor mit BamHI und EcoRI heraus-  
geschnitten, isoliert und in den geschnittenen Vektor unter  
Standardbedingungen ligiert. Entsprechend positive Klone, welche  
die vollständige *codA*-RNAi Kasse-  
20 *codA*sense-STLS1-*codA*anti-35S-T), werden durch analytischen  
Restriktionsverdau und Sequenzierung identifiziert.

Der Transfer der *codA*-RNAi Kasse-  
ette in den pSUN-1 Vektor  
(SEQ ID NO: 56) erfolgt unter Verwendung der die Kasse-  
25 flankierenden Restriktionsschnitt-stellen SacI und KpnI. Der  
entstandene Vektor pSUN1-*codA*-RNAi (siehe Fig. 7; SEQ ID NO: 57)  
wird zur Transformation von transgenen *A.thaliana* Pflanzen  
verwendet, die ein aktives *codA* Gen exprimieren (s.u.). Der  
Pflanzenexpressions-Vektor pSUN-1 ist im Rahmen des erfindungs-  
30 gemäßen Verfahrens besonders geeignet, da er keinen weiteren  
positiven Selektionsmarker trägt.

Der entstandene Vektor pSUN1-*codA*-RNAi ermöglicht die konsti-  
tutive Expression einer artifiziellen *codA*-dsRNA Variante,  
35 bestehend aus zwei identischen Nukleinsäureelementen, die durch  
ein Intron getrennt, in invertierter Form zueinander vorliegen.  
In Folge der Transkription dieser artifiziellen *codA*-dsRNA  
Variante kommt es aufgrund der Komplementarität der invertierten  
Nukleinsäureelemente zur Ausbildung eines doppelsträngigen RNA  
40 Moleküls. Das Vorhandensein dieses Moleküls induziert die Unter-  
drückung der Expression (RNA Akkumulation) des *codA* Gens mittels  
„double strand RNA interference“.

Beispiel 4: Herstellung transgener *Arabidopsis thaliana* Pflanzen

Transgene *Arabidopsis thaliana* Pflanzen, die als Markerprotein das *codA* Gen aus *E.coli* transgen exprimieren ("A.thaliana-  
5 [codA]"), wurden hergestellt wie beschrieben (Kirik et al. (2000) EMBO J 19(20):5562-6).

Die A.thaliana-[codA] Pflanzen werden mit einem *Agrobacterium tumefaciens* Stamm (GV3101 [pMP90]) auf Grundlage einer modifi-  
10 zierten Vakuuminfiltrationsmethode transformiert (Clough S & Bent A (1998) Plant J 16(6):735-43; Bechtold N et al. (1993) CR Acad Sci Paris 1144(2):204-212). Die verwendeten *Agrobacterium tumefaciens* Zellen werden im Vorfeld mit dem beschriebenen DNA-Konstrukt (pSUN1-codA-RNAi) transformiert. Auf diese Art werden  
15 doppelt-transgene A.thaliana-[codA] Pflanzen erzeugt, die unter Kontrolle des konstitutiven Nitrilase1-Promotors eine artifizielle *codA*-doppelsträngige RNA exprimieren. Als Folge des durch die Anwesenheit dieser artifiziellen *codA*-dsRNA induzierten dsRNAi-Effektes wird die Expression des *codA* Gens unterdrückt.  
20 Diese doppelt-transgenen Pflanzen können aufgrund ihrer wieder-gewonnen Fähigkeit, in Anwesenheit von 5-Fluorocytosin im Kultur-medium zu wachsen, identifiziert werden.

Samen der Primärtransformanten werden auf Grundlage der wieder-  
25 gewonnenen Fähigkeit in Anwesenheit von 5-Fluorocytosin zu wachsen selektioniert. Zu diesem Zweck werden die T1 Samen der Primärtransformanten auf Selektionsmedium ausgelegt welches 200 µg/ml 5-Fluorocytosin enthält. Diese Selektionsplatten werden unter Langtagbedingungen (16 Std. Licht, 21°C/8 Std. Dunkel, 18°C) inkubi-  
30 biert. Keimlinge, die sich in Anwesenheit von 5-Fluorocytosin normal entwickeln, werden nach 7 Tagen separiert und auf neue Selektionsplatten transferiert. Diese Platten werden bei unveränderten Bedingungen für weitere 14 inkubiert. Anschließend werden die resistenten Keimlinge in Erde pikiert und unter Kurz-  
35 tagbedingungen (8 Std. Licht, 21°C/16 Std. Dunkel, 18°C) kultiviert. Nach 14 Tagen werden die jungen Pflanzen in das Gewächshaus transferiert und unter Kurztagbedingungen kultiviert.



## 84

Beispiel 5: Herstellung eines Pflanzentransformationsvektors enthaltend eine Expressions-kassette zur Expression einer doppelsträngigen codA RNA und eines pflanzlichen Selektions-markers

5

In den pSUN1-codA-RNAi (siehe Fig. 7; SEQ ID NO: 57) wird ein pflanzlicher Selektionsmarker bestehend aus einer mutierten Variante des A.thaliana Als-Gens kodierend für die Acetolactat - Synthase unter Kontrolle des Promotor des A.thaliana Actin-2 Gens

10 (Meagher RB &amp; Williamson RE (1994) The plant cytoskeleton.

In The Plant Cytoskeleton (Meyerowitz, E. & Somerville, C., eds), pp. 1049-1084. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) und des Terminators der Octopin-Synthase (GIELLEN J et al. (1984) EMBO J 3:835-846) eingefügt (At.Act.-2-At.Als-

15 R-ocsT).

Dazu wird der Vektor pSUN1-codA-RNAi zunächst mit dem Restriktionsenzym Pvu II linearisiert. In diesen linearisierten Vektor erfolgt anschließend unter Standardbedingungen die Ligation eines

20 linearen DNA Fragmentes mit stumpfen Enden, kodierend für eine mutierte Variante der Acetolactat-Synthase (Als-R-Gen). Dieses

DNA Fragment wurde im Vorfeld der Ligation mit dem Restriktionsenzym KpnI verdaut und die überhängenden Enden durch eine Behandlung mit der Pwo DNA-Polymerase (Roche) gemäß der Herstellervor-

25 gaben in stumpfe Enden überführt. Diese mutierte Variante des Als Gens aus A.thaliana kann nicht durch Herbizide des Imidazolinon-Typs inhibiert werden. Durch Expression dieses mutierten A.tAls-R Gens erlangen die Pflanzen die Fähigkeit, in Anwesenheit des Herbizides Pursuit™ zu wachsen. Entsprechend positive Klone

30 (pSUN1-codA-RNAi-At.Act.-2-At.Als-R-ocsT; SEQ ID NO: 57) werden durch analytischen Restriktionsverdau und Sequenzierung identifiziert.

Der entstandene Vektor ermöglicht die konstitutive Expression

35 einer artifiziellen codA RNA Variante (bestehend aus zwei identischen Nukleinsäureelementen, die durch ein Intron getrennt, in invertierter Form zueinander vorliegen) und einer mutierten Variante des A.thaliana Als-Gens. In Folge der Transkription dieser artifiziellen codA RNA Variante kommt es aufgrund der

40 Komplementarität der invertierten Nukleinsäureelemente zur Ausbildung eines doppelsträngigen RNA Moleküls. Das Vorhandensein dieses Moleküls induziert die Unterdrückung der Expression (RNA Akkumulation) des codA Gens mittels "double strand RNA interference". Die Expression des Als-R Gens vermittelt den Pflanzen

45 die Fähigkeit, in Anwesenheit von Herbiziden des imidazolinon-Typs zu wachsen.

**Beispiel 6: Herstellung transgener Arabidopsis thaliana Pflanzen**

Transgene Arabidopsis thaliana Pflanzen, die als Markerprotein das codA Gen aus E.coli exprimieren ("A.thaliana-[codA]"), wurden wie beschrieben (Kirik et al.(2000) EMBO J 19(20):5562-6) hergestellt.

Die A.thaliana-[codA] Pflanzen werden mit einem Agrobacterium tumefaciens Stamm (GV3101 [pMP90]) auf Grundlage einer modifizierten Vakuuminfiltrationsmethode transformiert (Clough S & Bent A (1998) Plant J 16(6):735-43; Bechtold N et al. (1993) CR Acad Sci Paris 1144(2):204-212). Die verwendeten Agrobacterium tumefaciens Zellen werden im Vorfeld mit dem beschriebenen DNA-Konstrukt (pSUN1-codA-RNAi-At.Act.-2-At.Als-R-ocsT; SEQ ID NO: 57) transformiert. Auf diese Art werden doppeltransgene A.thaliana-[codA] Pflanzen erzeugt, die zusätzlich unter Kontrolle des konstitutiven Nitrilase1-Promotors eine artifizielle codA doppelsträngige RNA und eine Herbizid insensitive Variante des Als-Gens (Als-R) exprimieren (A.thaliana-[codA]-[codA-RNAi-At.Act.-2-At.Als-R-ocsT]). Als Folge des durch die Anwesenheit dieser artifiziellen codA-dsRNA induzierten dsRNAi-Effektes wird die Expression des codA Gens unterdrückt. Diese doppeltransgenen Pflanzen können aufgrund ihrer wiedergewonnen Fähigkeit, in Anwesenheit von 5-Fluorocytosin im Kulturmedium zu wachsen, identifiziert werden. Zusätzlich können positiv transformierte Pflanzen aufgrund ihrer Fähigkeit, in Anwesenheit des Herbizides Pursuit im Kulturmedium zu wachsen, selektioniert werden.

Zum Zweck der Selektion werden daher die T1 Samen der Primärtransformanten auf Selektionsmedium ausgelegt welches 100 µg/ml 5-Fluorocytosin enthält. Diese Selektionsplatten werden unter Langtagbedingungen (16 Std. Licht, 21°C/8 Std. Dunkel, 18°C) inkubiert. Keimlinge, die sich in Anwesenheit von 5-Fluorocytosin normal entwickeln, werden nach 28 Tagen separiert und auf neue Selektionsplatten transferiert. Diese Platten werden bei unveränderten Bedingungen für weitere 14 Tage inkubiert. Anschließend werden die resistenten Keimlinge in Erde pikiert und unter Kurztagbedingungen (8 std. Licht, 21°C/16 Std. Dunkel, 18°C) kultiviert. Nach weiteren 14 Tagen werden die jungen Pflanzen in das Gewächshaus transferiert und unter Kurztagbedingungen kultiviert.

Zusätzlich können Samen der Primärtransformanten - aufgrund ihrer Fähigkeit in Anwesenheit des Herbizides Pursuit™ zu wachsen - selektioniert werden. Es ist weiterhin möglich, eine Doppelselektion unter Verwendung des Herbizides Pursuit™ und 5-Fluorocytosin im Selektionsmedium durchzuführen. Zu diesem Zweck werden die T1 Samen der Primärtransformanten auf Selektionsmedium ausge-

## 86

legt welches das Herbizid Pursuit™ in einer Konzentration von 100 nM enthält (im Falle der Doppelselektion ist ebenfalls 100 µg/ml 5-Fluorocytosin enthalten). Diese Selektionsplatten werden unter Langtagbedingungen (16 Std. Licht, 21°C/8 Std. Dunkel, 18°C) inkubiert.

Keimlinge, die sich in Anwesenheit von Pursuit™ (Pursiut™ und 5-Fluorocytosin) normal entwickeln, werden nach 28 Tagen separiert und auf neue Selektionsplatten transferiert. Diese Platten werden bei unveränderten Bedingungen für weitere 14 Tage inkubiert. Anschließend werden die resistenten Keimlinge in Erde pikiert und unter Kurztagbedingungen (8 std. Licht, 21°C/16 Std. Dunkel, 18°C) kultiviert. Nach 14 Tagen werden die jungen Pflanzen in das Gewächshaus transferiert und unter Kurztagbedingungen kultiviert.

Beispiel 7: Analyse der unter Verwendung von 5-Fluorocytosin und/oder Pursuit selektionierten doppeltransgenen A.thaliana Pflanzen (A.thaliana-[codA]-[codA-RNAi-At.Act.-2-At.Als-R-ocsT])

Die Integration der T-DNA Region des zur Transformation verwendeten Vektors pSUN1-codA-RNAi-A.tAls-R in die genomische DNA der Ausgangspflanze (A.thaliana-[codA]) und der Verlust der codA spezifischen mRNA in diesen transgenen Pflanzen (A.thaliana-[codA]-[codA-RNAi- At.Act.-2-At.Als-R-ocsT]) kann unter Anwendung von Southern-Analysen und PCR Techniken bzw. Northern-Analysen nachgewiesen werden.

Um diese Analysen durchzuführen wird gesamt RNA und DNA (unter Verwendung des RNeasy Maxi Kit (RNA) bzw. Dneasy Plant Maxi Kit (genomische DNA) gemäß Herstellerangaben von Qiagen) aus Blattgewebe der transgenen Pflanzen und geeigneter Kontrollen isoliert. Bei den PCR Analysen kann die genomische DNA direkt als Grundlage (Template) der PCR verwendet werden. Die gesamt-RNA wird im Vorfeld der PCR in cDNA umgeschrieben. Die cDNA Synthese erfolgt unter Verwendung der Reversen Transkriptase Superscript II (Invitrogen) gemäß der Herstellerangaben.

40

45

## 87

- Beispiel 8: Nachweis der Reduktion der *codA* RNA Fließgleichgewichtsmenge in den positiv selektionierten doppelt-transgenen Pflanzen (*A.thaliana* [*codA*]-[*codA*-RNAi-*At.Act.-2-At.Als-R-ocsT*]) im Vergleich zu den zur Transformation verwendeten Ausgangspflanzen (*A.thaliana* [*codA*]) durch cDNA Synthese mit anschließender PCR Amplifikation.

PCR Amplifikation der *codA* spezifischen cDNA:

- 10 Die cDNA des *codA* Gens (ACCESSION S56903) kann unter Verwendung eines sense spezifischen Primers (*codA*5'C-term SEQ ID NO: 69) und eines antisense spezifischen Primers (*codA*3'C-term SEQ ID NO: 70) amplifiziert werden. Zu wählende PCR Bedingungen sind die folgenden:

- 15 Die PCR erfolgte in einem 50 µl Reaktionsansatz in dem enthalten ist:

- 2µl (200ng)cDNA aus *A.thaliana* -[*codA*] bzw. *A.thaliana* [*codA*]-[*codA*-RNAi-*At.Act.-2-At.Als-R-ocsT*]-Pflanzen
- 20 - 0,2 mM dATP, dTTP, dGTP, dCTP
- 1,5 mM Mg(OAc)<sub>2</sub>
- 5 µg Rinderserum-Albumin
- 40 pmol *codA*5'C-term SEQ ID NO: 69
- 40 pmol *codA*3'C-term SEQ ID NO: 70
- 25 - 15 µl 3,3x rTth DNA Polymerase XLPuffer (PE Applied Biosystems)
- 5U rTth DNA Polymerase XL (PE Applied Biosystems)

- 30 Die PCR wurde unter folgenden Zyklus Bedingungen durchgeführt:
- Schritt 1: 5 Minuten 94°C (Denaturierung)
- Schritt 2: 3 Sekunden 94°C
- Schritt 3: 1 Minute 56°C (Annealing)
- Schritt 4: 2 Minuten 72°C (Elongation)
- 30 Wiederholungen der Schritte 2 bis 4
- 35 Schritt 5: 10 Minuten 72°C (Post-Elongation)
- Schritt 6: 4°C (Warteschleife)

- In den positiv selektionierten Pflanzen ist die mRNA Fließgleichgewichtsmenge des *codA* Gens und die daraus resultierende
- 40 Menge an CODA-Proteins derart reduziert, dass eine quantitative Umsetzung von 5-Fluorocytosin zu 5-Fluorouracil nicht mehr erfolgen kann. Die Folge ist, dass diese Pflanzen (im Gegensatz zu den nicht transformierten Pflanzen) in Anwesenheit von 5-Fluorocytosin wachsen können. Somit wird nachgewiesen, dass transgene
- 45 Pflanzen aufgrund des angewendeten Prinzips der Verhinderung der Expression eines negativen Selektionsmarkers identifiziert werden

können.

- Beispiel 9: Nachweis der für die *codA*-RNAi kodierenden DNA unter Verwendung genomischer DNA der positiv selektionierten doppeltransgenen Pflanzen (*A.thaliana* [codA]-[codA-RNAi- At.Act.-2-At.Als-R-ocsT])

- Das *codA*-RNAi Transgen kann unter Verwendung eines *codA* spezifischen Primers (z.B. *codA*5'HindIII SEQ ID NO: 50) und eines 35s-Terminator spezifischen Primers amplifiziert (35sT 5' Primer SEQ ID NO: 71) werden. Durch Verwendung dieser Primerkombination kann spezifisch nur die für das *codA* RNAi Konstrukt codierende DNA nachgewiesen werden, da das *codA* Gen, welches in den zur Transformation verwendeten Ausgangspflanzen (*A.thaliana* [codA]) bereits vorhanden war, durch den nos-Terminator flankiert wird.

Zu wählende PCR Bedingungen sind die folgenden:

Die PCR erfolgte in einem 50 µl Reaktionsansatz in dem enthalten ist:

- 2 µl (200ng) genomische DNA aus den *A.thaliana* [codA]-[codA-RNAi-At.Act.-2-At.Als-R-ocsT]-Pflanzen
- 0,2 mM dATP, dTTP, dGTP, dCTP
- 1,5 mM Mg(OAc)<sub>2</sub>
- 5 µg Rinderserum-Albumin
- 40 pmol *codA* spezifischer sense Primer (SEQ ID NO: 50, 53 oder 69)
- 40 pmol 35sT 5' Primer SEQ ID NO: 71
- 15 µl 3,3x rTth DNA Polymerase XLPuffer (PE Applied Biosystems)
- 5U rTth DNA Polymerase XL (PE Applied Biosystems)

Die PCR wurde unter folgenden Zyklus Bedingungen durchgeführt:

- Schritt 1: 5 Minuten 94°C (Denaturierung)
- Schritt 2: 3 Sekunden 94°C
- Schritt 3: 1 Minute 56°C (Annealing)
- Schritt 4: 2 Minuten 72°C (Elongation)
- 30 Wiederholungen der Schritte 2-4
- Schritt 5: 10 Minuten 72°C (Post-Elongation)
- Schritt 6: 4°C (Warteschleife)

- In den positiv selektionierten Pflanzen kann auf diesem Weg die Integration des *codA*-RNAi DNA Konstruktes in die chromosomale DNA der zur Transformation verwendeten Ausgangspflanzen nachgewiesen werden. Somit wird nachgewiesen, dass transgene Pflanzen aufgrund des angewendeten Prinzips der Verhinderung der Expression eines negativen Selektionsmarkers identifiziert werden können.

- Beispiel 10: Nachweis der Reduktion der codA RNA Fleißgleichgewichtsmenge in den positiv selektionierten doppelt-transgenen Pflanzen (A.thaliana [codA]-[codA-RNAi-At.Act.-2-At.Als-R-ocsT]) im Vergleich zu den zur Transformation verwendeten Ausgangspflanzen (A.thaliana [codA]) durch Northern-Analyse.

Gelelektrophoretische Auftrennung von RNA:

- 10 Es wird pro RNA-Agarosegel 3 g Agar in 150 ml H<sub>2</sub>O (f.c. 1,5 % (w/v)) in der Mikrowelle gelöst und auf 60°C abgekühlt. Durch Zugabe von 20 ml 10x MEN (0,2 M MOPS, 50 mM Natriumacetat, 10 mM EDTA) und 30 ml Formaldehyd (f.c. 2,2 M) tritt weitere Abkühlung ein, so dass die gut gemischte Lösung zügig gegossen werden muss.
- 15 Formaldehyd verhindert die Bildung von Sekundärstrukturen in der RNA, deshalb ist die Laufgeschwindigkeit dem Molekulargewicht annähernd proportional (LEHRBACH H et al. (1977) Biochem J 16: 4743-4751). Die RNA-Proben werden vor Auftrag auf das Gel in folgendem Ansatz denaturiert: 20 µl RNA (1-2 µg/µl), 5 µl 10x MEN-Puffer, 6 µl Formaldehyd, 20 µl Formamid.
- 20

- Der Ansatz wird gemischt und 10 Minuten bei 65°C inkubiert. Nach Zugabe von 1/10 Volumen Probenpuffer und 1 µl Ethidiumbromid (10 mg/ml) wird die Probe aufgetragen. Die Gelelektrophorese erfolgt auf horizontalen Gelen in 1x MEN bei 120 V für zwei bis
- 25 drei Stunden. Nach der Elektrophorese wird das Gel unter UV-Licht unter Zuhilfenahme eines Lineals zur späteren Fragmentlängenbestimmung photographiert. Es folgt der RNA-Blot auf eine Nylonmembran gemäß der Angaben in: SAMBROOK J et al. Molecular cloning: A laboratory manual. Cold Spring Harbor, New York,
- 30 Cold Spring Harbor Laboratory Press. 1989.

Radioaktive Markierung von DNA-Fragmenten und Northern Hybridisierung

- 35 Zur Markierung des codA cDNA Fragmentes (codARNai-sense SEQ ID No: 49) kann z.B. der von Roche Diagnostics vertriebene High Prime Kit verwendet werden. Der "High prime" Kit basiert auf der von Feinberg und Vogelstein ursprünglich beschriebenen "random primed" Methode zur DNA Markierung. Zur Markierung werden ca.
- 40 25 ng DNA in 9-11 µl H<sub>2</sub>O für 10 min. bei 95°C denaturiert. Nach kurzer Inkubation auf Eis werden 4 µl High Prime Lösung (enthält eine random primer Mischung, 4 Einheiten Klenow Polymerase und jeweils 0,125 mM dATP, dTTP und dGTP in einem Reaktions-Puffer mit 50 % Glycerinanteil) und 3-5 µl [ $\alpha^{32}$ P]dCTP (30-50 µCi) hinzugegeben.
- 45 Der Ansatz wird für mindestens 10 min bei 37°C inkubiert und das nicht eingebaute dCTP anschließend von der nunmehr radioaktiv markierten DNA durch Gelfiltration über eine Sephadex

G-50-Säule getrennt. Anschließend wird das Fragment 10 min bei 95°C denaturiert und bis zur Verwendung auf Eis gehalten. Es werden folgende Hybridisierungs- und Vorinkubationspuffer benutzt:

5

Hypo Hybond  
250 mM Natriumphosphat-Puffer pH 7,2  
1 mM EDTA  
7 % SDS (g/v)

10 250 mM NaCl

10 µg/ml ssDNA  
5 % Polyethylenglykol (PEG) 6000  
40 % Formamid

- 15 Bei Verwendung von Hypo Hybond beträgt die Hybridisierungstemperatur 42°C, die Hybridisierungsdauer 16-24 Std. Zum Waschen der RNA-Filter werden drei verschiedene Lösungen verwendet 2 x SSC (300 mM NaCl; 30 mM NaCitrat) + 0,1% SDS, 1 x SSC + 0,1 % SDS und 0,1 x SSC + 0,1 % SDS. Die Länge und Intensität des Waschens
- 20 richtet sich nach der Stärke der gebundenen Aktivität. Im Anschluss an das Waschen werden die Filter in Plastikfolie eingeschweißt und ein Röntgenfilm (X-Omat, Kodak) bei -70 °C über Nacht exponiert. Die Signalstärke auf den Röntgenfilmen ist ein Maß für die Menge der codA mRNA Moleküle in der auf den Membranen
- 25 gebundenen gesamt RNA. In den positiv selektionierten Pflanzen kann somit die Reduktion der codA mRNA im Vergleich zu den zur Transformation verwendeten Ausgangspflanzen nachgewiesen werden.

- In den positiv selektionierten Pflanzen ist die mRNA Fließ-
- 30 gleichgewichtsmenge des codA Gens und die resultierende Menge an gebildeten CODA-Proteins derart reduziert, dass eine quantitative Umsetzung von 5-Fluorocytosin zu 5-Fluorouracil nicht mehr erfolgen kann. Die Folge ist, dass diese Pflanzen (im Gegensatz zu den nicht transformierten Pflanzen) in Anwesenheit von
- 35 5-Fluorocytosin wachsen können. Somit wird nachgewiesen, dass transgene Pflanzen aufgrund des angewendeten Prinzips der Verhinderung der Expression eines negativen Selektionsmarkers identifiziert werden können.

#### 40 Beispiel 11: Zusammenfassung der Ergebnisse der "Negativ-Negativ" Selektion

- Die Transformation der codA-transgenen Arabidopsis Pflanzen mit dem codA-dsRNA Konstrukt (pSUN1-codA-RNAi-At.Act.-2-At.Als-
- 45 R-ocsT; SEQ ID NO: 57) führt sowohl bei der Einfach-Selektion (nut mit 5-Fluorocytosin) als auch bei der Doppel-Selektion (Pursiut™ und 5-Fluorocytosin) zu einer signifikant erhöhten

## 91

Anzahl von doppelt-transgene Pflanzen, die das RNAi-Konstrukt erfolgreich in das Genom integriert haben (jeweils im Vergleich zu nicht transformierten Pflanzen). Die Analyse mittels PCR (s.o.) bestätigt bei der Mehrzahl der so generierten Pflanzen 5 den doppelt transgenen Status. Damit kann erfolgreich die Praktikabilität der vorliegenden Erfindung d.h. die Nutzbarkeit der Repression eines negativen Markers zur positiven Selektion (quasi eine "negativ-negativ" Selektion) gezeigt werden.

10

15

20

25

30

35

40

45



## Patentansprüche

1. Verfahren zur Herstellung transformierter pflanzlicher Zellen  
oder Organismen umfassend nachfolgende Schritte:
  - a) Transformation einer Population pflanzlicher Zellen,  
wobei die Zellen besagter Population mindestens ein  
Markerprotein enthalten, das für besagte Population  
direkt oder indirekt einen toxischen Effekt bewirken  
kann, mit mindestens einer zu insertierenden Nuklein-  
säuresequenz in Kombination mit mindestens einer Ver-  
bindung befähigt zur Verminderung der Expression, Menge,  
Aktivität und/oder Funktion mindestens eines Marker-  
proteins, und
  - b) Selektion von transformierten pflanzlichen Zellen, die in  
ihrem Genom besagte Nukleinsäuresequenz aufweisen und die  
infolge der Wirkung besagter Verbindung gegenüber nicht-  
transformierten Zellen einen Wachstumsvorteil haben,  
aus besagter Population pflanzlicher Zellen, wobei die  
Selektion unter Bedingungen durchgeführt wird, bei denen  
das Markerprotein seinen toxischen Effekt auf die nicht-  
transformierten Zellen ausüben kann.
2. Verfahren nach Anspruch 1, wobei das Markerprotein in der  
Lage ist, eine für besagte Population pflanzlicher Zellen  
nicht-toxische Substanz X in eine für besagte Population  
toxische Substanz Y direkt oder indirekt umzusetzen,  
umfassend nachfolgende Schritte:
  - a) Transformation der Population pflanzlicher Zellen mit  
mindestens einer zu insertierenden Nukleinsäuresequenz in  
Kombination mit mindestens einer Verbindung befähigt zur  
Verminderung der Expression, Menge, Aktivität und/oder  
Funktion mindestens eines Markerproteins, und
  - b) Behandlung besagter Population pflanzlicher Zellen mit  
der Substanz X in einer Konzentration, die infolge der  
Umsetzung durch das Markerprotein einen für nicht-trans-  
formierte Zellen toxischen Effekt bedingt, und
  - c) Selektion von transformierten pflanzlichen Zellen, die in  
ihrem Genom besagte Nukleinsäuresequenz aufweisen und die  
infolge der Wirkung besagter Verbindung gegenüber nicht-  
transformierten Zellen einen Wachstumsvorteil haben,  
aus besagter Population pflanzlicher Zellen, wobei die

Selektion unter Bedingungen durchgeführt wird, bei denen das Markerprotein seinen toxischen Effekt auf die nicht-transformierten Zellen ausüben kann.

- 5 3. Verfahren nach Anspruch 2, wobei es sich bei der nicht-toxischen Substanz X um eine Substanz handelt, die natürlicherweise in pflanzlichen Zellen oder Organismen nicht oder nur in Konzentration vorkommt, die im wesentlichen keinen toxischen Effekt bewirken können.
- 10 4. Verfahren nach Anspruch 2 oder 3, wobei es sich bei der Substanz X um eine Substanz handelt ausgewählt aus der Gruppe bestehend aus Pro-Herbiziden, Pro-Antibiotika, Nukleosidanaloga, 5-Fluorocytosin, Auxinamidverbindungen, Naphthalacetamid, Dihaloalkanen, Acyclovir, Ganciclovir, 1,2-Deoxy-2-fluoro- $\beta$ -D-arabinofuranosil-5-iodouracil, 6-Thioxanthin, 15 Allopurinol, 6-Methylpurindeoxyribonukleosid, 4-Aminopyrazolopyrimidin, 2-Amino-4-methoxy-butansäure, 5-(Tri-fluoromethyl)thioribose und Allylalkohol.
- 20 5. Verfahren nach einem der Ansprüche 1 bis 4, wobei das Markerprotein ausgewählt ist aus der Gruppe bestehend aus Cytosindeaminasen, Cytochrom P-450 Enzymen, Indolessigsäurehydro-lasen, Haloalkandehalogenasen, Thymidinkinasen, Guaninphos-25 phoribosyltransferasen, Hypoxanthinphosphoribosyltrans-ferasen, Xanthinguaninphosphoribosyltransferasen, Purin-nukleosidphosphorylasen, Phosphonatmonoesterhydrolasen, Indolacetamidsynthasen, Indolacetamidhydrolasen, Adenin-phosphoribosyltransferasen, Methoxinindehydrogenasen, 30 Rhizobitoxinsynthasen, 5-Methylthioribosekinasen und Alkohol-dehydrogenasen.
- 35 6. Verfahren nach einem der Ansprüche 1 bis 5, wobei das Markerprotein kodiert wird durch
- 35 a) eine Sequenz beschrieben durch die GenBank Accession-Nummer S56903, M32238, NC003308, AE009419, AB016260, NC002147 M26950, J02224, V00470, V00467, U10247, M13422, X00221, M60917, U44852, M61151, AF039169, AB025110, 40 AF212863, AC079674, X77943, M12196, AF172282, X04049 oder AF253472
- 45 b) eine Sequenz gemäß SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46 oder 48

7. Verfahren nach einem der Ansprüche 1 bis 6, wobei die Verbindung befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion mindestens eines Markerproteins mindestens eine Nukleinsäuresequenz, Ribonukleinsäuresequenz, doppelsträngige Ribonukleinsäuresequenz, antisense-Ribonukleinsäuresequenz, Expressionskassette oder Polypeptidsequenz umfasst.
8. Verfahren nach einem der Ansprüche 1 bis 7, wobei die Verbindung befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion mindestens eines Markerproteins ein DNA-Konstrukt ist, welches umfasst
- a) mindestens eine Expressionskassette geeignet zur Expression einer Ribonukleinsäuresequenz und/oder gegebenenfalls eines Proteins, wobei besagte Nukleinsäuresequenz und/oder Protein in der Lage ist, die Expression, Menge, Aktivität und/oder Funktion des Markerproteins zu vermindern, oder
- b) mindestens eine Sequenz, die eine teilweise oder vollständige Deletion oder Inversion der Sequenz kodierend für besagtes Markerprotein bewirkt und so eine Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins ermöglicht, sowie - gegebenenfalls - weitere Funktionselemente, die besagte Deletion oder Inversion erleichtern und/oder fördern, oder
- c) mindestens eine Sequenz, die eine Insertion in die Sequenz kodierend für das Markerprotein bewirkt und so eine Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins ermöglicht, sowie - gegebenenfalls - weitere Funktionselemente, die besagte Insertion erleichtern und/oder fördern.
9. Verfahren nach einem der Ansprüche 1 bis 8, wobei die Verminderung der Expression, Menge, Aktivität und/oder Funktion des Markerproteins durch mindestens eines der nachfolgenden Verfahren realisiert wird
- a) Einbringen mindestens einer doppelsträngigen Markerprotein Ribonukleinsäuresequenz oder einer deren Expression gewährleistenden Expressionskassette oder Expressionskassetten.

## 95

- b) Einbringen mindestens einer Markerprotein antisense-Ribonukleinsäuresequenzen oder einer deren Expression gewährleistenden Expressionskassette.
- 5 c) Einbringen mindestens einer Markerprotein antisense-Ribonukleinsäuresequenz kombiniert mit einem Ribozym oder einer deren Expression gewährleistenden Expressionskassette oder Expressionskassetten.
- 10 d) Einbringen mindestens einer Markerprotein sense-Ribonukleinsäuresequenzen zur Induktion einer Kosuppression oder einer deren Expression gewährleistenden Expressionskassette.
- 15 e) Einbringen mindestens eines DNA-oder Protein-bindenden Faktors gegen ein Markerprotein-Gen, -RNA oder -Protein oder einer dessen Expression gewährleistenden Expressionskassette.
- 20 f) Einbringen mindestens einer den Markerprotein RNA-Abbau bewirkenden viralen Nukleinsäuresequenz oder einer deren Expression gewährleistenden Expressionskassette.
- 25 g) Einbringen mindestens eines Konstruktes zur Erzeugung einer Insertion, Deletion, Inversion oder Mutation in einem Markerprotein-Gen.
10. Verfahren nach einem der Ansprüche 1 bis 9, wobei die Inaktivierung des Markerprotein-Gens durch das Einbringen einer sequenzspezifischen Rekombinase realisiert wird.
- 30 11. Verfahren nach einem der Ansprüche 1 bis 9, wobei die Inaktivierung des Markerprotein-Gens durch das Einbringen eines sequenzspezifischen Enzyms geeignet zur Induktion von DNA-Doppelstrangbrüchen realisiert wird.
- 35 12. Verfahren nach einem der Ansprüche 1 bis 11, wobei die Inaktivierung des Markerprotein-Gens durch Induktion einer intramolekularen oder intermolekularen homologen Rekombination realisiert wird.
- 40 13. Verfahren nach Anspruch 12, wobei die homologe Rekombination durch Einwirkung eines sequenzspezifischen Enzyms geeignet zur Induktion von DNA-Doppelstrangbrüchen gefördert wird.
- 45

14. Verfahren nach einem der Ansprüche 1 bis 13, wobei die Inaktivierung des Markerproteins durch Integration einer DNA-Sequenz in ein Markerprotein-Gen realisiert wird, umfassend nachfolgende Schritte:
- 5
- i) Einbringen eines Insertionskonstruktes und mindestens eines Enzyms geeignet zur Induktion von DNA-Doppelstrangbrüchen an einer Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen in oder in der Nähe
- 10 des Markerprotein-Gens, und
- ii) Induktion von DNA-Doppelstrangbrüchen an den Erkennungssequenzen zur gezielten Induktion von DNA-Doppelstrangbrüchen in oder in der Nähe des Markerprotein-Gens, und
- 15
- iii) Insertion des Insertionskonstruktes in das Markerprotein-Gen, wobei die Funktionalität des Markerprotein-Gens und - bevorzugt - die Funktionalität der Erkennungssequenz zur gezielten Induktion von DNA-Doppelstrangbrüchen
- 20 inaktiviert wird, so dass besagte Erkennungssequenz nicht mehr durch das Enzym geeignet zur Induktion von DNA-Doppelstrangbrüchen geschnitten werden kann, und
- iv) Selektion von Pflanzen oder pflanzlichen Zellen, bei
- 25 denen das Insertionskonstrukt in das Markerproteingen inseriert wurde.
15. Verfahren nach einem der Ansprüche 11, 13 oder 14, wobei das sequenzspezifische Enzym geeignet zur Induktion von DNA-Doppelstrangbrüchen eine Homing-Endonuklease ist.
- 30
16. Verfahren nach einem der Ansprüche 1 bis 15, wobei zusammen mit der zu insertierenden Nukleinsäuresequenz eine Sequenz kodierend für eine Resistenz gegen mindestens ein Toxin, Antibiotikum oder Herbizid eingebracht wird, und die
- 35 Selektion zusätzlich unter Einsatz des Toxins, Antibiotikums oder Herbizids erfolgt.
17. Verfahren nach einem der Ansprüche 1 bis 16, wobei die
- 40 in das Genom der pflanzlichen Zelle oder des pflanzlichen Organismus zu insertierende Nukleinsäuresequenz mindestens eine Expressionskassette umfasst, wobei besagte Expressionskassette unter Kontrolle eines in pflanzlichen Zellen oder pflanzlichen Organismen funktionellen Promotors eine RNA
- 45 und/oder ein Protein exprimieren kann, welche nicht die Ver-

minderung der Expression, Menge, Aktivität und/oder Funktion eines Markerproteins bewirken.

18. Verfahren nach einem der Ansprüche 1 bis 17, wobei die  
5 pflanzliche Zelle Teil eines pflanzlichen Organismus oder eines davon abgeleiteten Gewebes, Teils, Organs, Zellkultur oder Vermehrungsmaterials ist.
19. Verfahren nach einem der Ansprüche 1 bis 18 zur Herstellung  
10 transformierter pflanzlicher Zellen oder Organismen, , umfassend nachfolgende Schritte:
  - a) Transformation einer Population pflanzlicher Zellen,  
15 welche mindestens ein nicht-endogenes (bevorzugt nicht-pflanzliches) Markerprotein umfasst, das in der Lage ist, eine für besagte Population pflanzlicher Zellen nicht-toxische Substanz X in eine für besagte Population toxische Substanz Y direkt oder indirekt umzusetzen, mit  
20 mindestens einer zu insertierenden Nukleinsäuresequenz in Kombination mit mindestens einer Nukleinsäuresequenz kodierend für eine Ribonukleinsäuresequenz befähigt zur Verminderung der Expression, Menge, Aktivität und/oder Funktion besagten Markerproteins, und
  - 25 b) Behandlung besagter Population pflanzlicher Zellen mit der Substanz X in einer Konzentration, die infolge der Umsetzung durch das Markerprotein einen für nicht-transformierte Zellen toxischen Effekt bedingt, und
  - 30 c) Selektion von transformierten pflanzlichen Zellen (und/oder Populationen pflanzlicher Zellen wie pflanzlichen Geweben oder Pflanzen), die in ihrem Genom besagte Nukleinsäuresequenz aufweisen und die infolge der Wirkung  
35 besagter Verbindung gegenüber nicht-transformierten Zellen einen Wachstumsvorteil haben, aus besagter Population pflanzlicher Zellen, wobei die Selektion unter Bedingungen durchgeführt wird, bei denen das Markerprotein seinen toxischen Effekt auf die nicht-transformierten Zellen ausüben kann, und
  - 40 d) Regeneration von fertilen Pflanzen, und
  - e) Auskreuzung der für das Markerprotein kodierenden Nukleinsäuresequenz und Isolation von fertilen Pflanzen,  
45 die in ihrem Genom besagte Nukleinsäuresequenz aber nicht

mehr die für das Markerprotein kodierende Sequenz aufweisen.

20. Aminosäuresequenzen kodieren für eine pflanzliche 5-Methylthioribosekinase, dadurch gekennzeichnet, dass besagte Aminosäuresequenz mindestens eine Sequenz enthält ausgewählt aus der Gruppe bestehend aus SEQ ID NO: 60, 62, 64, 66 oder 68.
21. Nukleinsäuresequenz kodieren für eine pflanzliche 5-Methylthioribosekinase, dadurch gekennzeichnet, dass besagte Nukleinsäuresequenz mindestens eine Sequenz enthält ausgewählt aus der Gruppe bestehend aus SEQ ID NO: 59, 61, 63, 65 oder 67.
22. Doppelsträngiges RNA-Molekül umfassend
- a) einen "sense"-RNA-Strang umfassend mindestens eine Ribonukleotidsequenz, die im wesentlichen identisch ist zu mindestens einem Teil des "sense"-RNA-Transkriptes einer Nukleinsäuresequenz kodierend für ein Markerprotein, und
- b) einen "antisense"-RNA-Strang, der zu dem RNA-sense-Strang unter a) im wesentlichen - bevorzugt vollständig - komplementären ist.
23. Doppelsträngiges RNA-Molekül nach Anspruch 22, wobei das Markerprotein wie in einem der Ansprüche 2 bis 6 definiert ist.
24. Doppelsträngiges RNA-Molekül nach einem der Ansprüche 22 oder 23, wobei "sense"-RNA-Strang und "antisense"-RNA-Strang kovalent in Form eines invertierten Repeats miteinander verbunden sind.
25. Transgene Expressionskassette enthaltend in funktioneller Verknüpfung mit einem in pflanzlichen Organismen funktionellen Promotor eine Nukleinsäuresequenz kodierend für ein doppelsträngiges RNA-Molekül gemäß einem der Ansprüche 22 bis 24.
26. Transgene Expressionskassette enthaltend eine Nukleinsäuresequenz kodierend für zumindest einen Teil eines Markerproteins, wobei besagte Nukleinsäuresequenz mit einem in pflanzlichen Organismen funktionellen Promotor in antisense-Orientierung funktionell verknüpft ist.

27. Transgene Expressionskassette nach Anspruch 26, wobei das Markerprotein wie in einem der Ansprüche 2 bis 6 definiert ist.
- 5 28. Transgener Vektor enthaltend eine transgene Expressionskassette gemäß einem der Ansprüche 25 bis 27.
- 10 29. Transgener pflanzlicher Organismus enthaltend ein doppelsträngiges RNA-Molekül gemäß einem der Ansprüche 22 bis 24, eine transgene Expressionskassette gemäß einem der Ansprüche 25 bis 27 oder einen transgenen Vektor gemäß Anspruch 28.
- 15 30. Transgener pflanzlicher Organismus nach Anspruch 29, ausgewählt aus der Gruppe der Pflanzen bestehend aus Weizen, Hafer, Hirse, Gerste, Roggen, Mais, Reis, Buchweizen, Sorghum, Triticale, Dinkel, Leinsamen, Zuckerrohr, Raps, Kresse, Arabidopsis, Kohlarten, Soja, Alfalfa, Erbse, Bohnengewächsen, Erdnuss, Kartoffel, Tabak, Tomate, Aubergine, 20 Paprika, Sonnenblume, Tagetes, Salat, Calendula, Melone, Kürbis und Zucchini.
- 25 31. Gewebe, Organ, Teil, Zelle, Zellkultur oder Vermehrungsgut abgeleitet von einem transgenen pflanzlichen Organismus gemäß einem der Ansprüche 29 oder 30.

30

35

40

45



1/11

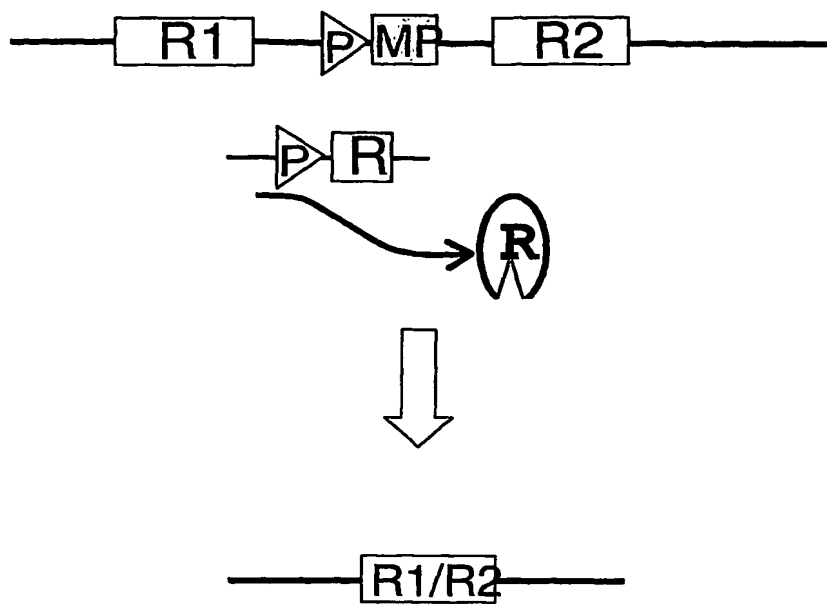


Fig. 1

2/11

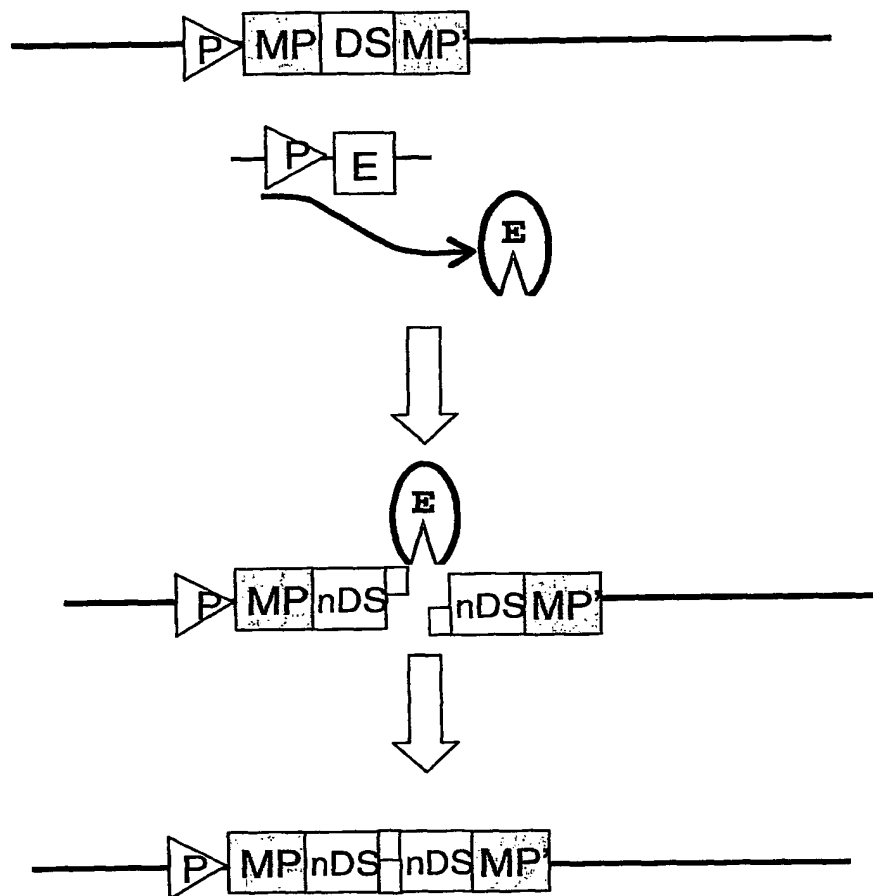


Fig. 2-A

3/11

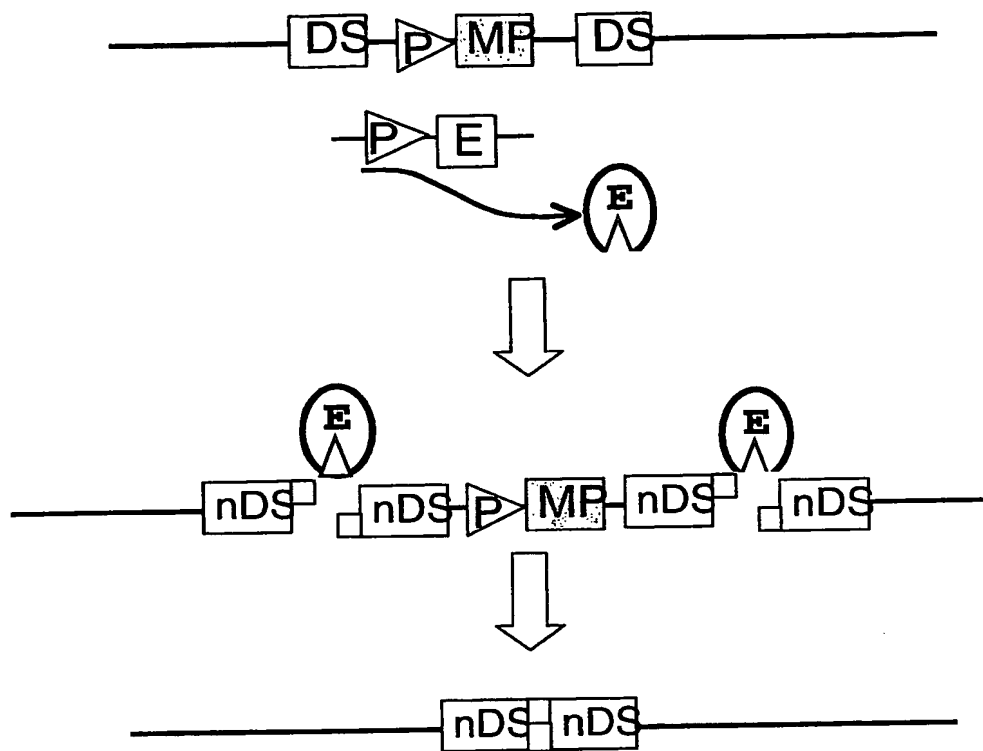


Fig. 2-B

4/11

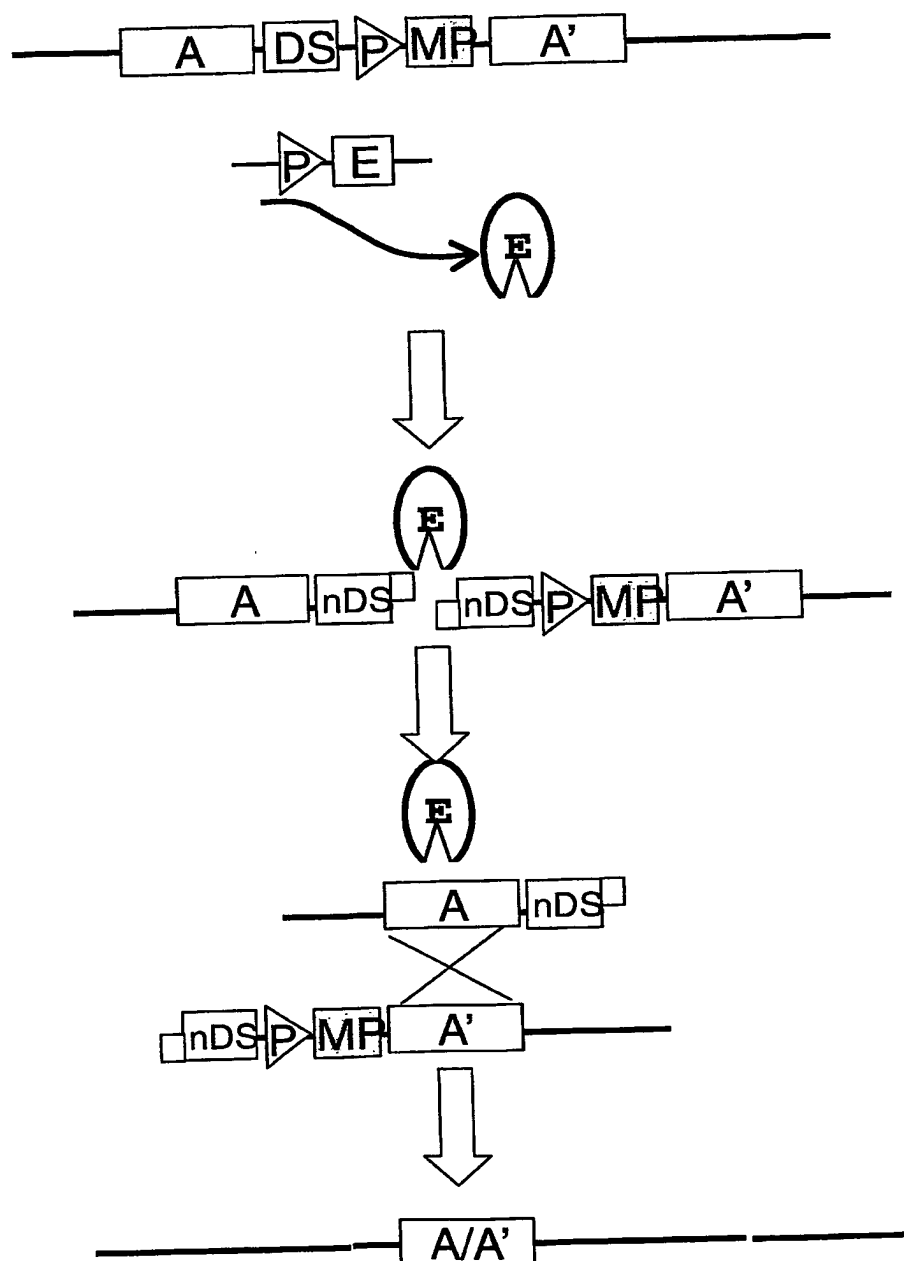


Fig. 3

5/11

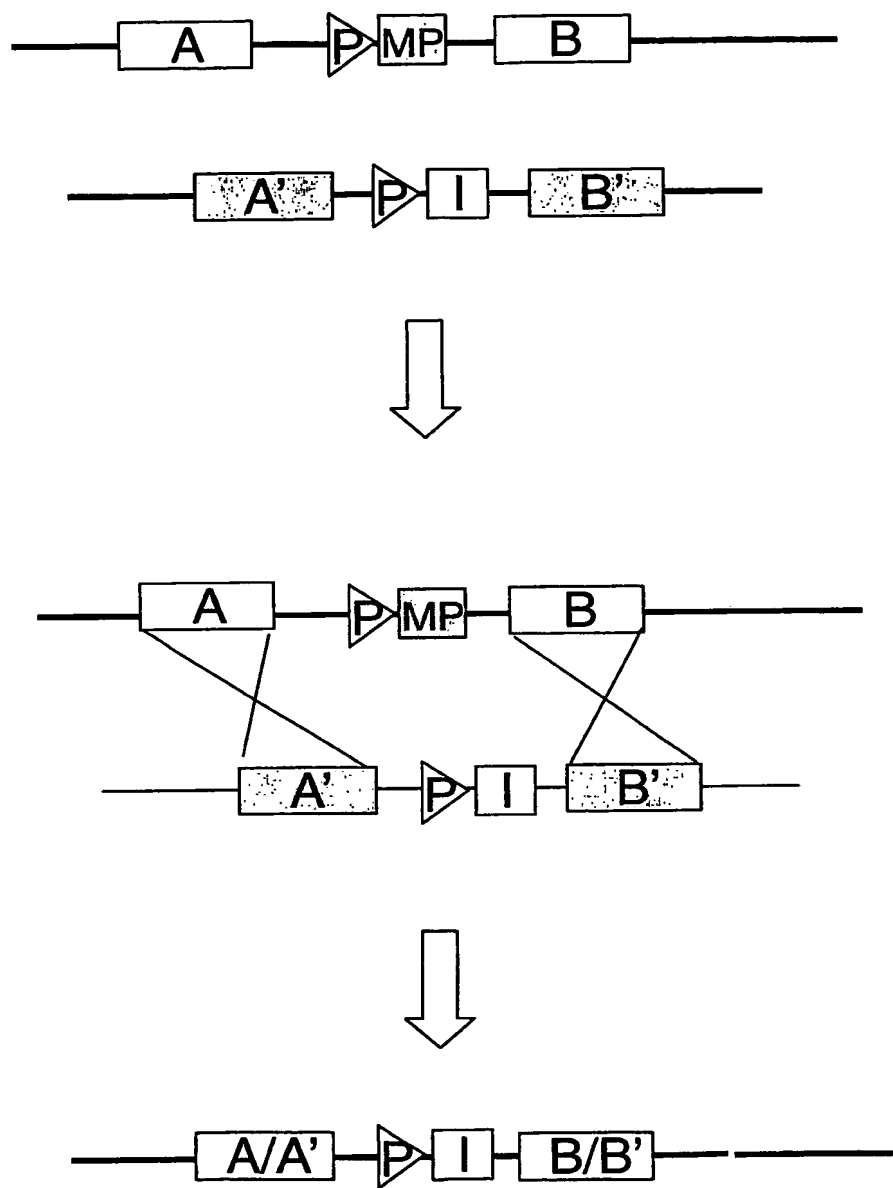


Fig. 4

6/11

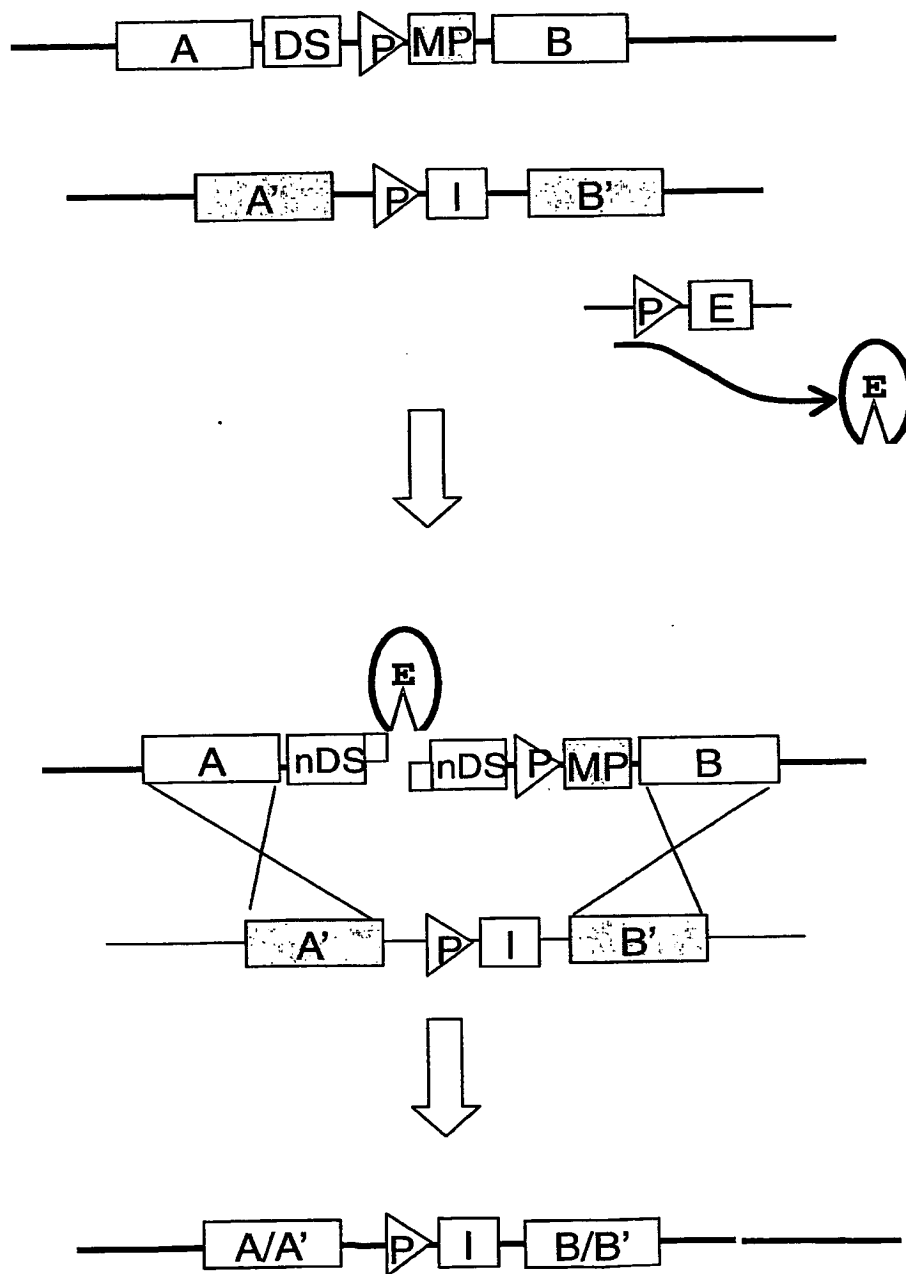


Fig. 5

7/11

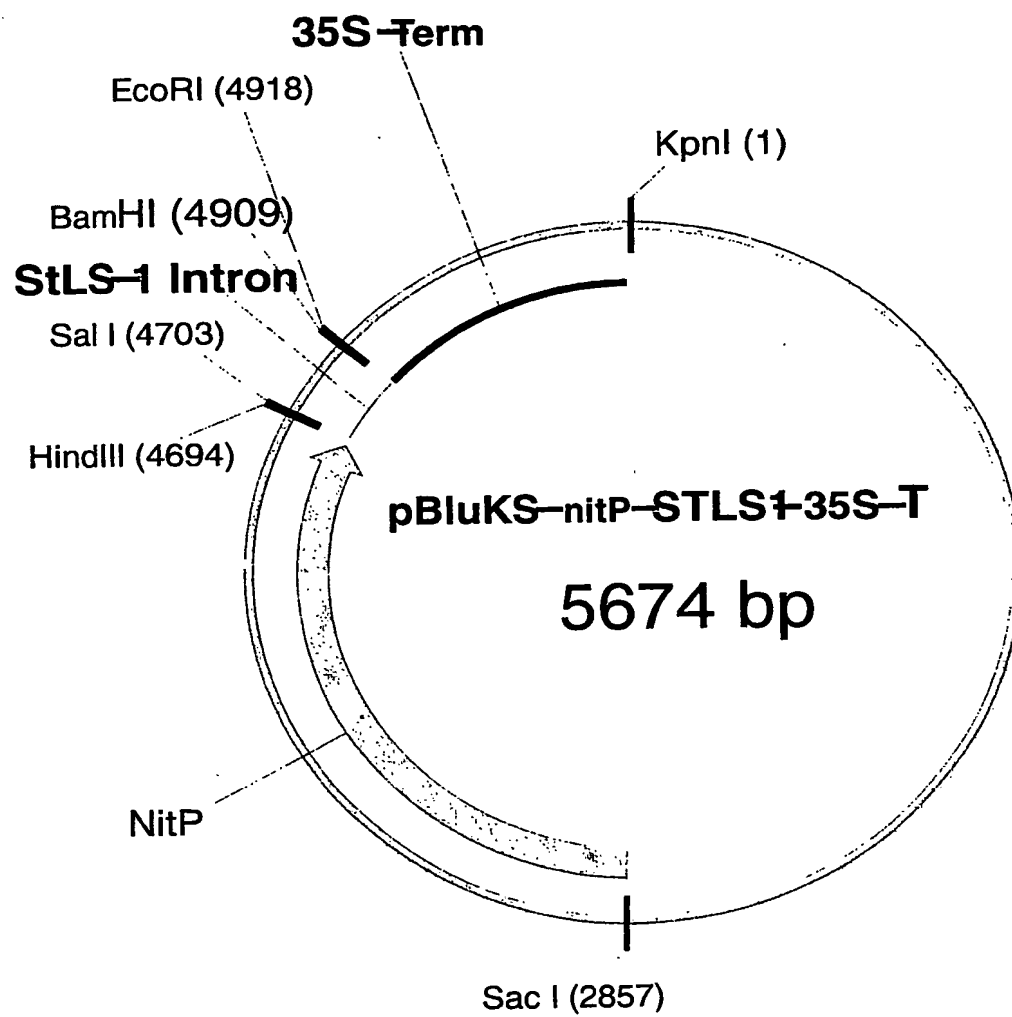


Fig. 6

8/11

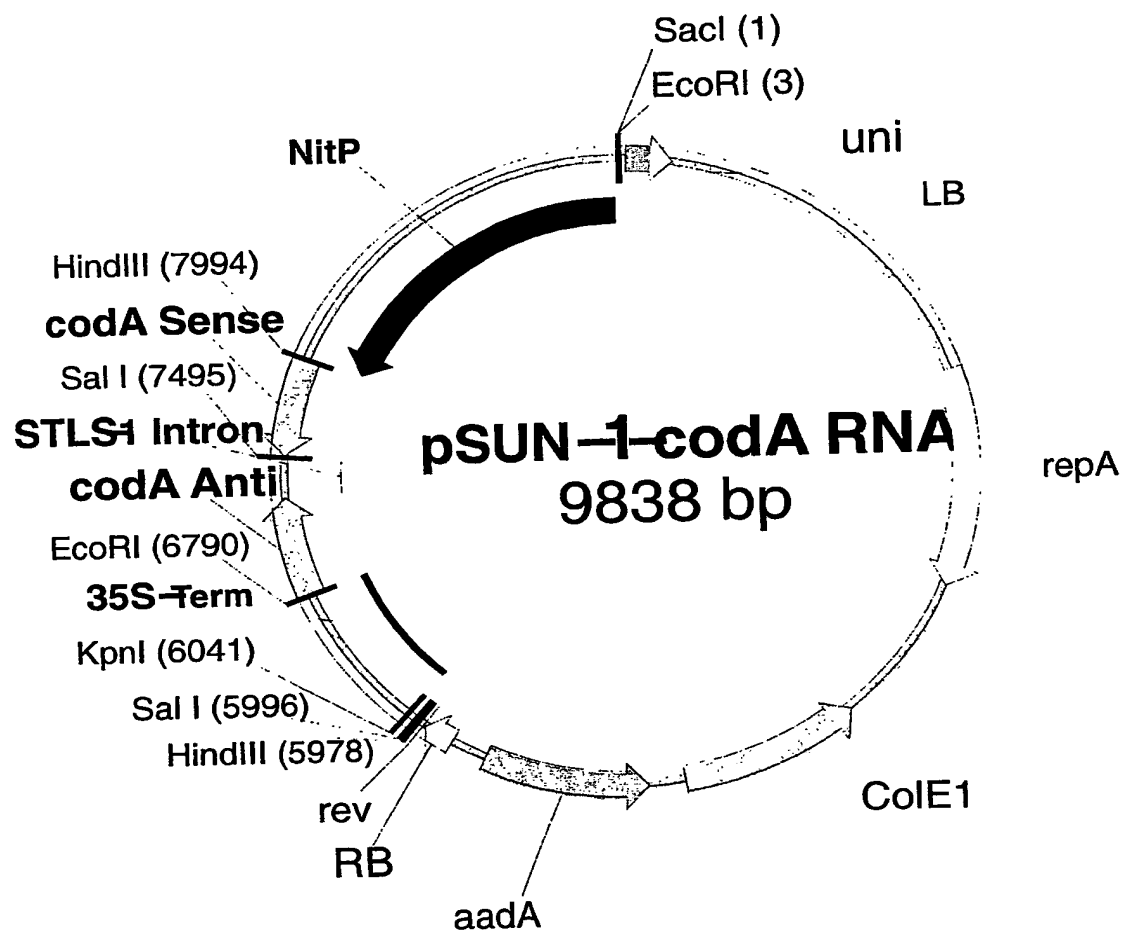
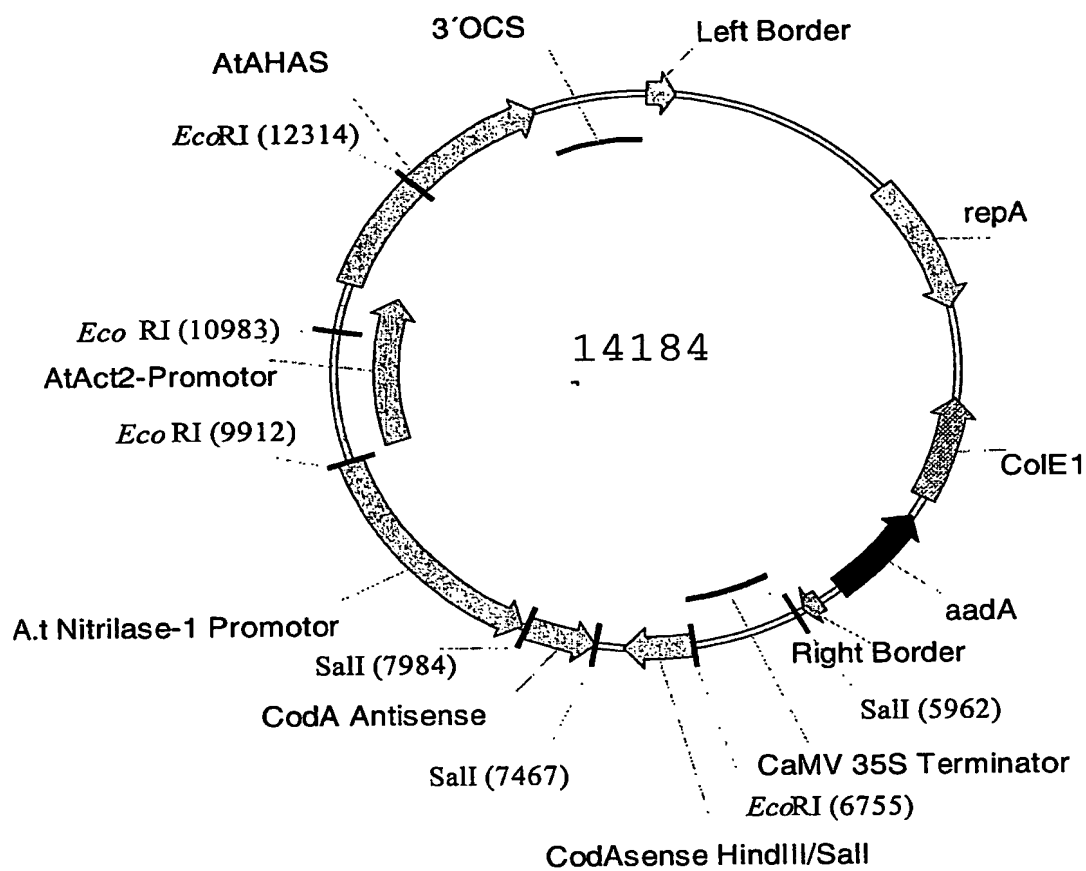


Fig. 7



9/11



pSUN1-codA-RNAi-At.Act.-2-At.A/s-R-ocsT

Fig. 8

10/11

	1	50
Klebsiella pneumoniae	(1) -----MSQYHTFTAHDVAAYAQQ	
Clostridium tetani.	(1) -----MSRFDShFRMETEDAILYAKE	
Zea mays	(1) ARALLSSPLAGASPCQASAMAAEEEQGFRPLDESSLLAYIKATPALAS	
A.thaliana	(1) -----MSFEEFTPLNEKSLVDYIKSTPALSS	
Brassica napus-2	(1) -----VDDFVLRKEMSDFEFKPLNEKSLVEYIKATPALSS	
Soy-1	(1) -----	
Oryza sativa-1	(1) -----	
Consensus	(1) -----L V A-----	
	51	100
Klebsiella pneumoniae	(19) FAGIDNPSELVSAQEVGDGNLNLVFKVFDRQGVSRRAIVKQALPYVRCVGE	
Clostridium tetani.	(22) KLGIFDEHAKLQAEIIGDGNINIVFKVWDVNTKKSUIKHADIFLRSSGR	
Zea mays	(51) RLGGGGSLLDSIEIKEVGDGNLNFVYIVQSEAGA--IVVKQALPYVRCVGD	
A.thaliana	(27) KIGADKSDDDLVIKEVGDGNLNFVFIIVGSSGS--LVIKQALPYIRCIGE	
Brassica napus -2	(37) RLGDKY--DDLVIKEVGDGNLNFVFIIVGSTGS--LVIKQALPYIRCIGE	
Soy -1	(1) -----	
Oryza sativa -1	(1) -----	
Consensus	(51) KLG D L EVGDGNLNFV V G LVIKQALPYIRCIGE	
	101	150
Klebsiella pneumoniae	(69) SWPLTLDRARLEAQTIVAHYQHSPQHTVKIHHFDP ELAVMV MEDLS-DHR	
Clostridium tetani.	(72) --ELDVDRNR IEAEV LMLQ GILAPGLVPKVYKYDSVMCNLSMEDIS-DHR	
Zea mays	(99) SWPMT RERAYFEASTLREHGRLCPEHTPEVYHFDR TSLMGMRYIEPPHI	
A.thaliana	(75) SWPMTKERAYFEATTLRKHG NLSPDHVPEVYHFDR TMALIGMRYLEPPHI	
Brassica napus -2	(83) SWPMTKERAYFEATTLRKHGGLSPDHVPEVYHFDR TMALIGMRYLEPPHI	
Soy -1	(1) -----IPEHVPEVYHFDR TMSLIGMRYLEPPHI	
sativa -1	(1) -----	
Consensus	(101) SWPMT ERA EA TL HG LSPDHVPEVYHFDR TMALIGMRYLEPPHI	
	151	200
Klebsiella pneumoniae	(118) IWRGELIANVYYPQAARQLGDYLAQVLFHTSDFY LHPHEKKAQVAQFIN-	
Clostridium tetani.	(119) NLRKEL LKRNTFPSFAEHITTFIVDTLLPTTDLVMDSGEKKDNVKKYIN-	
Zea mays	(149) ILRKGLVAGVEYPLLADHMSDYMAKTLFFTSLLYNNTTDHKNGVAKYSAN	
A.thaliana	(125) ILRKGLIAGIEYPFLADHMSDYMAKTLFFTSLLYHDTTEHRRAVTEFCGN	
Brassica napus -2	(133) ILRKG-----	
Soy -1	(29) ILIKGLIAGIEYPFLAEHMADFMAKTLFFTSLLFRSTADHKRDVAEFCGN	
Oryza sativa -1	(1) -----LLYNSTTDHKKGVAQYCDN	
Consensus	(151) ILRKGLIA I YP ADHM DYMA TLF TSLLY T DHK VA F N	
	201	250
Klebsiella pneumoniae	(167) PAMCEITEDLFFNDPYQIHERN--NYP AELEADVAALRDDAQLKLAVAAL	
Clostridium tetani.	(168) KDLCKISED LVFTEPFIDYKSRNTVLEENIEFVKRQLYEDKELILEAGKL	
Zea mays	(199) VEMCRLTEQVVFSDPYRVSKFNR-WTSPYLDKDAEAVREDELKLEVAGL	
A.thaliana	(175) VELCRLTEQVVFSDPYRVSTFNR-WTSPYLD DDAKAVREDSALKLEIAEL	
Brassica napus -2	(138) -----	
Soy -1	(79) VELCRLTEQVVFSDPYKVSQYNR-WTSPYLD RDAEAVREDNLLKLEVAEL	
Oryza sativa -1	(20) VEMCRLTEQVVFSDPYMLAKYNR-CTSPFLDNDA AAVREDAELKLEIAEL	
Consensus	(201) VELCRLTEQVVFSDPY VS FNR TSPYLD DA AVRED LKLEVA L	

Fig. 9a

11/11

251	300
Klebsiella pneumoniae	(215) KHRFFAHAEALLHGDIHSGSIFVAEGSLKAIDAEFGYFGPIGFDIGTAIG
Clostridium tetani.	(218) KNNFMNNSQALIHGDLHSGSIFVNEESTKILDPEFAFYGPIGYDLGNVIG
Zea mays	(248) KSMFIERAQALIHGDLHTGSMVTEVQLKSLIQNLGSMGPMGFDIGSLPW
A.thaliana	(224) KSMFCERAQALIHGDLHTGSMVMTQDSTQVIDPEFSFYGPMGFDIGAYLG
Brassica napus -2	(138) -----
Soy -1	(128) KSKFIES-----
Oryza sativa -1	(69) KSMFIERAQALLHGDLHTGSMVTPDSTQVIDPEFAFYGPMGYDIGAFLG
Consensus	(251) KS FIE AQALIHGDLHTGSI V S ID EFAFYGPMGFDIG IG
	301 350
Klebsiella pneumoniae	(265) NLLLNLCGLPGQLGIRDAAAAREQRLNDIHQLWTTFAERFQALAAEKTRD
Clostridium tetani.	(268) NLFFAWANAYVTEDGKEVEEFTIWIIEKTIENILELFKEKFIKKYKEIVTD
Zea mays	(298) KPDFGHTMHRMGMLIKRMIVRLTRMDLEDN-----
A.thaliana	(274) NLILAFFAQDGHATQENDRKEYKQWILRTIEQTWNLFNKRFIALWDQNKD
Brassica napus -2	(138) -----
Soy -1	(135) -----
Oryza sativa -1	(119) NLILAYFSQDGHADQANDRKAY-----
Consensus	(301) NL AY
	351 400
Klebsiella pneumoniae	(315) AALAYPGYASAFLLKKVWADAVGFCGSELIRRSVGLSHVADIDTIQDDAMR
Clostridium tetani.	(318) VMAKEEYYMNWYLHSILSDTAGQVGLIIRRVGDSKVLDTITSITDINKR
Zea mays	(328) -----
A.thaliana	(324) GPGEAYLADIYNNTEVLKFVQENYMRNLLHDSLGFGAAMIRRIVGVAHV
Brassica napus -2	(138) -----
Soy -1	(135) -----
Oryza sativa -1	(141) -----
Consensus	(351)
	401 447
Klebsiella pneumoniae	(365) HECLRHAITLGRALIVLAERIDSVDELLARVRQYS-----
Clostridium tetani.	(368) VKAERILILSAKTFIKNRHKIKTGKRYVEIFNSNMY-----
Zea mays	(328) -----
A.thaliana	(374) EDFESI EEDKRRRAICERSALEFAKMLLKERRKFKSIGEVVSAIQQS
Brassica napus -2	(138) -----
Soy -1	(135) -----
Oryza sativa -1	(141) -----
Consensus	(401)

Fig. 9b

## SEQUENZPROTOKOLL

&lt;110&gt; BASF Plant Science GmbH

&lt;120&gt; Neue Selektionssysteme

&lt;130&gt; PF53790-AT

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 71

&lt;170&gt; PatentIn Ver. 2.1

&lt;210&gt; 1

&lt;211&gt; 1284

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1281)

&lt;223&gt; coding for cytosine deaminase (codA)

&lt;400&gt; 1

gtg tcg aat aac gct tta caa aca att att aac gcc cgg tta cca ggc	48
Val Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly	
1 5 10 15	
gaa gag ggg ctg tgg cag att cat ctg cag gac gga aaa atc agc gcc	96
Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala	
20 25 30	
att gat gcg caa tcc ggc gtg atg ccc ata act gaa aac agc ctg gat	144
Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp	
35 40 45	
gcc gaa caa ggt tta gtt ata ccg ccg ttt gtg gag cca cat att cac	192
Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His	
50 55 60	
ctg gac acc acg caa acc gcc gga caa ccg aac tgg aat cag tcc ggc	240
Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly	
65 70 75 80	
acg ctg ttt gaa ggc att gaa cgc tgg gcc gag cgc aaa gcg tta tta	288
Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu	
85 90 95	
acc cat gac gat gtg aaa caa cgc gca tgg caa acg ctg aaa tgg cag	336
Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln	
100 105 110	
att gcc aac ggc att cag cat gtg cgt acc cat gtc gat gtt tcg gat	384
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp	
115 120 125	
gca acg cta act gcg ctg aaa gca atg ctg gaa gtg aag cag gaa gtc	432
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val	
130 135 140	
gcg ccg tgg att gat ctg caa atc gtc gcc ttc cct cag gaa ggg att	480
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile	
145 150 155 160	
ttg tcg tat ccc aac ggt gaa gcg ttg ctg gaa gag gcg tta cgc tta	528
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu	
165 170 175	

```

ggg gca gat gta gtg ggg gcg att ccg cat ttt gaa ttt acc cgt gaa 576
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
180 185 190

tac ggc gtg gag tcg ctg cat aaa acc ttc gcc ctg gcg caa aaa tac 624
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
195 200 205

gac cgt ctc atc gac gtt cac tgt gat gag atc gat gac gag cag tcg 672
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
210 215 220

cgc ttt gtc gaa acc gtt gct gcc ctg gcg cac cat gaa ggc atg ggc 720
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
225 230 235 240

gcg cga gtc acc gcc agc cac acc acg gca atg cac tcc tat aac ggg 768
Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
245 250 255

gcg tat acc tca cgc ctg ttc cgc ttg ctg aaa atg tcc ggt att aac 816
Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn
260 265 270

ttt gtc gcc aac ccg ctg gtc aat att cat ctg caa gga cgt ttc gat 864
Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp
275 280 285

acg tat cca aaa cgt cgc ggc atc acg cgc gtt aaa gag atg ctg gag 912
Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu
290 295 300

tcc ggc att aac gtc tgc ttt ggt cac gat gat gtc ttc gat ccg tgg 960
Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp
305 310 315 320

tat ccg ctg gga acg gcg aat atg ctg caa gtg ctg cat atg ggg ctg 1008
Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu
325 330 335

cat gtt tgc cag ttg atg ggc tac ggg cag att aac gat ggc ctg aat 1056
His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn
340 345 350

tta atc acc cac cac agc gca agg acg ttg aat ttg cag gat tac ggc 1104
Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly
355 360 365

att gcc gcc gga aac agc gcc aac ctg att atc ctg ccg gct gaa aat 1152
Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn
370 375 380

ggg ttt gat gcg ctg cgc cgt cag gtt ccg gta cgt tat tcg gta cgt 1200
Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg
385 390 395 400

ggc ggc aag gtg att gcc agc aca caa ccg gca caa acc acc gta tat 1248
Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr
405 410 415

ctg gag cag cca gaa gcc atc gat tac aaa cgt tga 1284
Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg
420 425

```

&lt;210&gt; 2

&lt;211&gt; 427

&lt;212&gt; PRT

&lt;213&gt; Escherichia coli

&lt;400&gt; 2

Val	Ser	Asn	Asn	Ala	Leu	Gln	Thr	Ile	Ile	Asn	Ala	Arg	Leu	Pro	Gly	1		5		10		15
Glu	Glu	Gly	Leu	Trp	Gln	Ile	His	Leu	Gln	Asp	Gly	Lys	Ile	Ser	Ala	20		25		30		
Ile	Asp	Ala	Gln	Ser	Gly	Val	Met	Pro	Ile	Thr	Glu	Asn	Ser	Leu	Asp	35		40		45		
Ala	Glu	Gln	Gly	Leu	Val	Ile	Pro	Pro	Phe	Val	Glu	Pro	His	Ile	His	50		55		60		
Leu	Asp	Thr	Thr	Gln	Thr	Ala	Gly	Gln	Pro	Asn	Trp	Asn	Gln	Ser	Gly	65		70		75		
Thr	Leu	Phe	Glu	Gly	Ile	Glu	Arg	Trp	Ala	Glu	Arg	Lys	Ala	Leu	Leu	85		90		95		
Thr	His	Asp	Asp	Val	Lys	Gln	Arg	Ala	Trp	Gln	Thr	Leu	Lys	Trp	Gln	100		105		110		
Ile	Ala	Asn	Gly	Ile	Gln	His	Val	Arg	Thr	His	Val	Asp	Val	Ser	Asp	115		120		125		
Ala	Thr	Leu	Thr	Ala	Leu	Lys	Ala	Met	Leu	Glu	Val	Lys	Gln	Glu	Val	130		135		140		
Ala	Pro	Trp	Ile	Asp	Leu	Gln	Ile	Val	Ala	Phe	Pro	Gln	Glu	Gly	Ile	145		150		155		
Leu	Ser	Tyr	Pro	Asn	Gly	Glu	Ala	Leu	Leu	Glu	Glu	Ala	Leu	Arg	Leu	165		170		175		
Gly	Ala	Asp	Val	Val	Gly	Ala	Ile	Pro	His	Phe	Glu	Phe	Thr	Arg	Glu	180		185		190		
Tyr	Gly	Val	Glu	Ser	Leu	His	Lys	Thr	Phe	Ala	Leu	Ala	Gln	Lys	Tyr	195		200		205		
Asp	Arg	Leu	Ile	Asp	Val	His	Cys	Asp	Glu	Ile	Asp	Asp	Glu	Gln	Ser	210		215		220		
Arg	Phe	Val	Glu	Thr	Val	Ala	Ala	Leu	Ala	His	His	Glu	Gly	Met	Gly	225		230		235		
Ala	Arg	Val	Thr	Ala	Ser	His	Thr	Thr	Ala	Met	His	Ser	Tyr	Asn	Gly	245		250		255		
Ala	Tyr	Thr	Ser	Arg	Leu	Phe	Arg	Leu	Leu	Lys	Met	Ser	Gly	Ile	Asn	260		265		270		
Phe	Val	Ala	Asn	Pro	Leu	Val	Asn	Ile	His	Leu	Gln	Gly	Arg	Phe	Asp	275		280		285		
Thr	Tyr	Pro	Lys	Arg	Arg	Gly	Ile	Thr	Arg	Val	Lys	Glu	Met	Leu	Glu	290		295		300		
Ser	Gly	Ile	Asn	Val	Cys	Phe	Gly	His	Asp	Asp	Val	Phe	Asp	Pro	Trp	305		310		315		
Tyr	Pro	Leu	Gly	Thr	Ala	Asn	Met	Leu	Gln	Val	Leu	His	Met	Gly	Leu	325		330		335		
His	Val	Cys	Gln	Leu	Met	Gly	Tyr	Gly	Gln	Ile	Asn	Asp	Gly	Leu	Asn	340		345		350		

att gcc aac ggc att cag cat gtg cgt acc cat gtc gat gtt tcg gat 384  
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp  
115 120 125

gca acg cta act gcg ctg aaa gca atg ctg gaa gtg aag cag gaa gtc	432
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val	
130 135 140	
gcg ccg tgg att gat ctg caa atc gtc gcc ttc cct cag gaa ggg att	480
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile	
145 150 155 160	
ttg tcg tat ccc aac ggt gaa gcg ttg ctg gaa gag gcg tta cgc tta	528
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu	
165 170 175	
ggg gca gat gta gtg ggg gcg att ccg cat ttt gaa ttt acc cgt gaa	576
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu	
180 185 190	
tac ggc gtg gag tgc ctg cat aaa acc ttc gcc ctg gcg caa aaa tac	624
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr	
195 200 205	
gac cgt ctc atc gac gtt cac tgt gat gag atc gat gac gag cag tgc	672
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser	
210 215 220	
cgc ttt gtc gaa acc gtt gct gcc ctg gcg cac cat gaa ggc atg ggc	720
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly	
225 230 235 240	
gcg cga gtc acc gcc agc cac acc acg gca atg cac tcc tat aac ggg	768
Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly	
245 250 255	
gcg tat acc tca cgc ctg ttc cgc ttg ctg aaa atg tcc ggt att aac	816
Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn	
260 265 270	
ttt gtc gcc aac ccg ctg gtc aat att cat ctg caa gga cgt ttc gat	864
Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp	
275 280 285	
acg tat cca aaa cgt cgc ggc atc acg cgc gtt aaa gag atg ctg gag	912
Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu	
290 295 300	
tcc ggc att aac gtc tgc ttt ggt cac gat gat gtc ttc gat ccg tgg	960
Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp	
305 310 315 320	
tat ccg ctg gga acg gcg aat atg ctg caa gtg ctg cat atg ggg ctg	1008
Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu	
325 330 335	
cat gtt tgc cag ttg atg ggc tac ggg cag att aac gat ggc ctg aat	1056
His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn	
340 345 350	
tta atc acc cac cac agc gca agg acg ttg aat ttg cag gat tac ggc	1104
Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly	
355 360 365	
att gcc gcc gga aac agc gcc aac ctg att atc ctg ccg gct gaa aat	1152
Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn	
370 375 380	
ggg ttt gat gcg ctg cgc cgt cag gtt ccg gta cgt tat tgc gta cgt	1200
Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg	
385 390 395 400	



ctg gag cag cca gaa gcc atc gat tac aaa cgt tga 1284  
Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg  
420 425

<223> Beschreibung der künstlichen Sequenz: coding for  
cytosine deaminase (codA)

Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp  
275 280 285

Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu  
 290 295 300  
 Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp  
 305 310 315 320  
 Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu  
 325 330 335  
 His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn  
 340 345 350  
 Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly  
 355 360 365  
 Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn  
 370 375 380  
 Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg  
 385 390 395 400  
 Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr  
 405 410 415  
 Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg  
 420 425

&lt;210&gt; 5

&lt;211&gt; 1221

&lt;212&gt; DNA

&lt;213&gt; Streptomyces griseolus

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1218)

&lt;223&gt; coding for cytochrome P450-Su1 (suaC)

&lt;400&gt; 5

atg acc gat acc gcc acg acg ccc cag acc acg gac gca ccc gcc ttc 48  
 Met Thr Asp Thr Ala Thr Thr Pro Gln Thr Thr Asp Ala Pro Ala Phe  
 1 5 10 15  
 ccg agc aac cgg agc tgt ccc tac cag tta ccg gac ggc tac gcc cag 96  
 Pro Ser Asn Arg Ser Cys Pro Tyr Gln Leu Pro Asp Gly Tyr Ala Gln  
 20 25 30  
 ctc cgg gac acc ccc ggc ccc ctg cac cgg gtg acg ctc tac gac ggc 144  
 Leu Arg Asp Thr Pro Gly Pro Leu His Arg Val Thr Leu Tyr Asp Gly  
 35 40 45  
 cgt cag gcg tgg gtg gtg acc aag cac gag gcc gcg cgc aaa ctg ctc 192  
 Arg Gln Ala Trp Val Val Thr Lys His Glu Ala Ala Arg Lys Leu Leu  
 50 55 60  
 ggc gac ccc cgg ctg tcc tcc aac cgg acg gac gac aac ttc ccc gcc 240  
 Gly Asp Pro Arg Leu Ser Ser Asn Arg Thr Asp Asp Asn Phe Pro Ala  
 65 70 75 80  
 acg tca ccg cgc ttc gag gcc gtc cgg gag agc ccg cag gcg ttc atc 288  
 Thr Ser Pro Arg Phe Glu Ala Val Arg Glu Ser Pro Gln Ala Phe Ile  
 85 90 95  
 ggc ctg gac ccg ccc gag cac ggc acc cgg cgg cgg atg acg atc agc 336  
 Gly Leu Asp Pro Pro Glu His Gly Thr Arg Arg Arg Met Thr Ile Ser  
 100 105 110

gag	ttc	acc	gtc	aag	cgg	atc	aag	ggc	atg	cgc	ccc	gag	gtc	gag	gag	384
Glu	Phe	Thr	Val	Lys	Arg	Ile	Lys	Gly	Met	Arg	Pro	Glu	Val	Glu	Glu	
		115						120				125				
gtg	gtg	cac	ggc	ttc	ctc	gac	gag	atg	ctg	gcc	gcc	ggc	ccg	acc	gcc	432
Val	Val	His	Gly	Phe	Leu	Asp	Glu	Met	Leu	Ala	Ala	Gly	Pro	Thr	Ala	
		130				135					140					
gac	ctg	gtc	agt	cag	ttc	gcg	ctg	ccg	gtg	ccc	tcc	atg	gtg	atc	tgc	480
Asp	Leu	Val	Ser	Gln	Phe	Ala	Leu	Pro	Val	Pro	Ser	Met	Val	Ile	Cys	
145					150					155					160	
cga	ctc	ctc	ggc	gtg	ccc	tac	gcc	gac	cac	gag	ttc	ttc	cag	gac	gcg	528
Arg	Leu	Leu	Gly	Val	Pro	Tyr	Ala	Asp	His	Glu	Phe	Phe	Gln	Asp	Ala	
				165					170					175		
agc	aag	cgg	ctg	gtg	cag	tcc	acg	gac	gcg	cag	agc	gcg	ctc	acc	gcg	576
Ser	Lys	Arg	Leu	Val	Gln	Ser	Thr	Asp	Ala	Gln	Ser	Ala	Leu	Thr	Ala	
			180					185					190			
cgg	aac	gac	ctc	gcg	ggt	tac	ctg	gac	ggc	ctc	atc	acc	cag	ttc	cag	624
Arg	Asn	Asp	Leu	Ala	Gly	Tyr	Leu	Asp	Gly	Leu	Ile	Thr	Gln	Phe	Gln	
		195					200					205				
acc	gaa	ccg	ggc	gcg	ggc	ctg	gtg	ggc	gct	ctg	gtc	gcc	gac	cag	ctg	672
Thr	Glu	Pro	Gly	Ala	Gly	Leu	Val	Gly	Ala	Leu	Val	Ala	Asp	Gln	Leu	
		210				215					220					
gcc	aac	ggc	gag	atc	gac	cgt	gag	gaa	ctg	atc	tcc	acc	gcg	atg	ctg	720
Ala	Asn	Gly	Glu	Ile	Asp	Arg	Glu	Glu	Leu	Ile	Ser	Thr	Ala	Met	Leu	
225					230					235					240	
ctc	ctc	atc	gcc	ggc	cac	gag	acc	acg	gcc	tcg	atg	acc	tcc	ctc	agc	768
Leu	Leu	Ile	Ala	Gly	His	Glu	Thr	Thr	Ala	Ser	Met	Thr	Ser	Leu	Ser	
				245					250					255		
gtg	atc	acc	ctg	ctg	gac	cac	ccc	gag	cag	tac	gcc	gcc	ctg	cgc	gcc	816
Val	Ile	Thr	Leu	Asp	His	Pro	Glu	Gln	Tyr	Ala	Ala	Ala	Leu	Arg	Ala	
			260				265						270			
gac	cgc	agc	ctc	gtg	ccc	ggc	gcg	gtg	gag	gaa	ctg	ctc	cgc	tac	ctc	864
Asp	Arg	Ser	Leu	Val	Pro	Gly	Ala	Val	Glu	Glu	Leu	Leu	Arg	Tyr	Leu	
		275					280					285				
gcc	atc	gcc	gac	atc	gcg	ggc	ggc	cgc	gtc	gcc	acg	gcg	gac	atc	gag	912
Ala	Ile	Ala	Asp	Ile	Ala	Gly	Gly	Arg	Val	Ala	Thr	Ala	Asp	Ile	Glu	
		290				295					300					
gtc	gag	ggg	cac	ctc	atc	cgg	gcc	ggc	gag	ggc	gtg	atc	gtc	gtc	aac	960
Val	Glu	Gly	His	Leu	Ile	Arg	Ala	Gly	Glu	Gly	Val	Ile	Val	Val	Asn	
305					310					315					320	
tcg	ata	gcc	aac	cgg	gac	ggc	acg	gtg	tac	gag	gac	ccg	gac	gcc	ctc	1008
Ser	Ile	Ala	Asn	Arg	Asp	Gly	Thr	Val	Tyr	Glu	Asp	Pro	Asp	Ala	Leu	
				325				330						335		
gac	atc	cac	cgc	tcc	gcg	cgc	cac	cac	ctc	gcc	ttc	ggc	ttc	ggc	gtg	1056
Asp	Ile	His	Arg	Ser	Ala	Arg	His	His	Leu	Ala	Phe	Gly	Phe	Gly	Val	
			340				345						350			
cac	cag	tgc	ctg	ggc	cag	aac	ctc	gcc	cgg	ctg	gag	ctg	gag	gtc	atc	1104
His	Gln	Cys	Leu	Gly	Gln	Asn	Leu	Ala	Arg	Leu	Glu	Leu	Glu	Val	Ile	
		355					360					365				
ctc	aac	gcc	ctc	atg	gac	cgc	gtc	ccg	acg	ctg	cga	ctg	gcc	gtc	ccc	1152
Leu	Asn	Ala	Leu	Met	Asp	Arg	Val	Pro	Thr	Leu	Arg	Leu	Ala	Val	Pro	
		370				375					380					

gtc gag cag ttg gtg ctg cgg ccg ggt acg acg atc cag ggc gtc aac 1200  
 Val Glu Gln Leu Val Leu Arg Pro Gly Thr Thr Ile Gln Gly Val Asn  
 385 390 395 400

gaa ctc ccg gtc acc tgg tga 1221  
 Glu Leu Pro Val Thr Trp  
 405

<210> 6

<211> 406

<212> PRT

<213> Streptomyces griseolus

<400> 6

Met Thr Asp Thr Ala Thr Thr Pro Gln Thr Thr Asp Ala Pro Ala Phe  
 1 5 10 15  
 Pro Ser Asn Arg Ser Cys Pro Tyr Gln Leu Pro Asp Gly Tyr Ala Gln  
 20 25 30  
 Leu Arg Asp Thr Pro Gly Pro Leu His Arg Val Thr Leu Tyr Asp Gly  
 35 40 45  
 Arg Gln Ala Trp Val Val Thr Lys His Glu Ala Ala Arg Lys Leu Leu  
 50 55 60  
 Gly Asp Pro Arg Leu Ser Ser Asn Arg Thr Asp Asp Asn Phe Pro Ala  
 65 70 75 80  
 Thr Ser Pro Arg Phe Glu Ala Val Arg Glu Ser Pro Gln Ala Phe Ile  
 85 90 95  
 Gly Leu Asp Pro Pro Glu His Gly Thr Arg Arg Arg Met Thr Ile Ser  
 100 105 110  
 Glu Phe Thr Val Lys Arg Ile Lys Gly Met Arg Pro Glu Val Glu Glu  
 115 120 125  
 Val Val His Gly Phe Leu Asp Glu Met Leu Ala Ala Gly Pro Thr Ala  
 130 135 140  
 Asp Leu Val Ser Gln Phe Ala Leu Pro Val Pro Ser Met Val Ile Cys  
 145 150 155 160  
 Arg Leu Leu Gly Val Pro Tyr Ala Asp His Glu Phe Phe Gln Asp Ala  
 165 170 175  
 Ser Lys Arg Leu Val Gln Ser Thr Asp Ala Gln Ser Ala Leu Thr Ala  
 180 185 190  
 Arg Asn Asp Leu Ala Gly Tyr Leu Asp Gly Leu Ile Thr Gln Phe Gln  
 195 200 205  
 Thr Glu Pro Gly Ala Gly Leu Val Gly Ala Leu Val Ala Asp Gln Leu  
 210 215 220  
 Ala Asn Gly Glu Ile Asp Arg Glu Glu Leu Ile Ser Thr Ala Met Leu  
 225 230 235 240  
 Leu Leu Ile Ala Gly His Glu Thr Thr Ala Ser Met Thr Ser Leu Ser  
 245 250 255  
 Val Ile Thr Leu Leu Asp His Pro Glu Gln Tyr Ala Ala Leu Arg Ala  
 260 265 270  
 Asp Arg Ser Leu Val Pro Gly Ala Val Glu Glu Leu Leu Arg Tyr Leu  
 275 280 285  
 Ala Ile Ala Asp Ile Ala Gly Gly Arg Val Ala Thr Ala Asp Ile Glu  
 290 295 300

Val Glu Gly His Leu Ile Arg Ala Gly Glu Gly Val Ile Val Val Asn  
 305 310 315 320  
 Ser Ile Ala Asn Arg Asp Gly Thr Val Tyr Glu Asp Pro Asp Ala Leu  
 325 330 335  
 Asp Ile His Arg Ser Ala Arg His His Leu Ala Phe Gly Phe Gly Val  
 340 345 350  
 His Gln Cys Leu Gly Gln Asn Leu Ala Arg Leu Glu Leu Glu Val Ile  
 355 360 365  
 Leu Asn Ala Leu Met Asp Arg Val Pro Thr Leu Arg Leu Ala Val Pro  
 370 375 380  
 Val Glu Gln Leu Val Leu Arg Pro Gly Thr Thr Ile Gln Gly Val Asn  
 385 390 395 400  
 Glu Leu Pro Val Thr Trp  
 405

&lt;210&gt; 7

&lt;211&gt; 1404

&lt;212&gt; DNA

&lt;213&gt; Agrobacterium tumefaciens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1401)

&lt;223&gt; coding for indole acetamide hydrolase (tms2)

&lt;400&gt; 7

atg gtg ccc att acc tcg tta gca caa acc cta gaa cgc ctg aga cgg	48
Met Val Pro Ile Thr Ser Leu Ala Gln Thr Leu Glu Arg Leu Arg Arg	
1 5 10 15	
aaa gac tac tcc tgc tta gaa cta gta gaa act ctg ata gcg cgt tgc	96
Lys Asp Tyr Ser Cys Leu Glu Leu Val Glu Thr Leu Ile Ala Arg Cys	
20 25 30	
caa gct gca aaa cca tta aat gcc ctt ctg gct aca gac tgg gat ggc	144
Gln Ala Ala Lys Pro Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Gly	
35 40 45	
ttg cgg cga agc gcc aaa aaa att gat cgt cat gga aac gcc gga tta	192
Leu Arg Arg Ser Ala Lys Lys Ile Asp Arg His Gly Asn Ala Gly Leu	
50 55 60	
ggt ctt tgc ggc att cca ctc tgt ttt aag gcg aac atc gcg acc ggc	240
Gly Leu Cys Gly Ile Pro Leu Cys Phe Lys Ala Asn Ile Ala Thr Gly	
65 70 75 80	
ata ttt cct aca agc gct gct act ccg gcg ctg ata aac cac ttg cca	288
Ile Phe Pro Thr Ser Ala Ala Thr Pro Ala Leu Ile Asn His Leu Pro	
85 90 95	
aag ata cca tcc cgc gtc gca gaa aga ctt ttt tca gct gga gca ctg	336
Lys Ile Pro Ser Arg Val Ala Glu Arg Leu Phe Ser Ala Gly Ala Leu	
100 105 110	
ccg ggt gcc tcg gga aac atg cat gag tta tcg ttt gga att acg agc	384
Pro Gly Ala Ser Gly Asn Met His Glu Leu Ser Phe Gly Ile Thr Ser	
115 120 125	
aac aac tat gcc acc ggt gcg gtg cgg aac ccg tgg aat cca agt ctg	432
Asn Asn Tyr Ala Thr Gly Ala Val Arg Asn Pro Trp Asn Pro Ser Leu	
130 135 140	

ata cca gga ggc tca agc ggt ggt gtg gct gct gcg gtg gca agc cga	480
Ile Pro Gly Gly Ser Ser Gly Gly Val Ala Ala Val Ala Ser Arg	
145 150 155 160	
ttg atg tta ggc ggc ata ggc acc gat acc ggt gca tct gtt cgc cta	528
Leu Met Leu Gly Gly Ile Gly Thr Asp Thr Gly Ala Ser Val Arg Leu	
165 170 175	
ccc gca gcc ctg tgt ggc gta gta gga ttt cga ccg acg ctt gct cga	576
Pro Ala Ala Leu Cys Gly Val Val Gly Phe Arg Pro Thr Leu Ala Arg	
180 185 190	
tat cca aga gat cgg ata ata ccg gtc agc ccc acc cgg gac acc gcc	624
Tyr Pro Arg Asp Arg Ile Ile Pro Val Ser Pro Thr Arg Asp Thr Ala	
195 200 205	
gga atc ata gcg cag tgc gta gcc gat gtt ata atc ctc gac cag gtg	672
Gly Ile Ile Ala Gln Cys Val Ala Asp Val Ile Ile Leu Asp Gln Val	
210 215 220	
att tcc gga cgg tcg gcg aaa att tca ccc atg ccg ctg aag ggg ctt	720
Ile Ser Gly Arg Ser Ala Lys Ile Ser Pro Met Pro Leu Lys Gly Leu	
225 230 235 240	
cgg atc ggc ctc ccc act acc tac ttt tac gat gac ctt gat gct gat	768
Arg Ile Gly Leu Pro Thr Thr Tyr Phe Tyr Asp Asp Leu Asp Ala Asp	
245 250 255	
gtg gcc ttc gca gct gaa acg acg att cgc ttg cta gcc aac aga ggc	816
Val Ala Phe Ala Ala Glu Thr Thr Ile Arg Leu Leu Ala Asn Arg Gly	
260 265 270	
gta acc ttt gtt gaa gcc gac atc ccc cac cta gag gaa ctg aat agt	864
Val Thr Phe Val Glu Ala Asp Ile Pro His Leu Glu Glu Leu Asn Ser	
275 280 285	
ggg gca agt ttg cca att gcg ctt tac gaa ttt cca cac gct cta aaa	912
Gly Ala Ser Leu Pro Ile Ala Leu Tyr Glu Phe Pro His Ala Leu Lys	
290 295 300	
aag tat ctc gac gat ttt gtg gga aca gtt tct ttt tct gac gtt atc	960
Lys Tyr Leu Asp Asp Phe Val Gly Thr Val Ser Phe Ser Asp Val Ile	
305 310 315 320	
aaa gga att cgt agc ccc gat gta gcg aac att gtc agt gcg caa att	1008
Lys Gly Ile Arg Ser Pro Asp Val Ala Asn Ile Val Ser Ala Gln Ile	
325 330 335	
gat ggg cat caa att tcc aac gat gaa tat gaa ctg gcg cgt caa tcc	1056
Asp Gly His Gln Ile Ser Asn Asp Glu Tyr Glu Leu Ala Arg Gln Ser	
340 345 350	
ttc agg cca agg ctc cag gcc act tat cgg aat tac ttc aga ctc tat	1104
Phe Arg Pro Arg Leu Gln Ala Thr Tyr Arg Asn Tyr Phe Arg Leu Tyr	
355 360 365	
cag tta gat gca atc ctt ttc cca act gca ccc tta gcg gcc aaa gcc	1152
Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Ala Ala Lys Ala	
370 375 380	
ata ggt cag gag tcg tca gtc atc cac aat ggc tca atg atg aac act	1200
Ile Gly Gln Glu Ser Ser Val Ile His Asn Gly Ser Met Met Asn Thr	
385 390 395 400	
ttc aag atc tac gtg cga aat gtg gac cca agc agc aac gca ggc cta	1248
Phe Lys Ile Tyr Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu	
405 410 415	

cct ggg ttg agc ctt cct gcc tgc ctt aca cct gat cgc ttg cct gtt 1296  
 Pro Gly Leu Ser Leu Pro Ala Cys Leu Thr Pro Asp Arg Leu Pro Val  
 420 425 430

gga atg gaa att gat gga tta gcg ggg tca gac cac cgt ctg tta gca 1344  
 Gly Met Glu Ile Asp Gly Leu Ala Gly Ser Asp His Arg Leu Leu Ala  
 435 440 445

atc ggg gca gca tta gaa aaa gcc ata aat ttt cct tcc ttt ccc gat 1392  
 Ile Gly Ala Ala Leu Glu Lys Ala Ile Asn Phe Pro Ser Phe Pro Asp  
 450 455 460

gct ttt aat tag 1404  
 Ala Phe Asn  
 465

<210> 8  
 <211> 467  
 <212> PRT  
 <213> *Agrobacterium tumefaciens*  
 <400> 8

Met Val Pro Ile Thr Ser Leu Ala Gln Thr Leu Glu Arg Leu Arg Arg  
 1 5 10 15

Lys Asp Tyr Ser Cys Leu Glu Leu Val Glu Thr Leu Ile Ala Arg Cys  
 20 25 30

Gln Ala Ala Lys Pro Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Gly  
 35 40 45

Leu Arg Arg Ser Ala Lys Lys Ile Asp Arg His Gly Asn Ala Gly Leu  
 50 55 60

Gly Leu Cys Gly Ile Pro Leu Cys Phe Lys Ala Asn Ile Ala Thr Gly  
 65 70 75 80

Ile Phe Pro Thr Ser Ala Ala Thr Pro Ala Leu Ile Asn His Leu Pro  
 85 90 95

Lys Ile Pro Ser Arg Val Ala Glu Arg Leu Phe Ser Ala Gly Ala Leu  
 100 105 110

Pro Gly Ala Ser Gly Asn Met His Glu Leu Ser Phe Gly Ile Thr Ser  
 115 120 125

Asn Asn Tyr Ala Thr Gly Ala Val Arg Asn Pro Trp Asn Pro Ser Leu  
 130 135 140

Ile Pro Gly Gly Ser Ser Gly Gly Val Ala Ala Ala Val Ala Ser Arg  
 145 150 155 160

Leu Met Leu Gly Gly Ile Gly Thr Asp Thr Gly Ala Ser Val Arg Leu  
 165 170 175

Pro Ala Ala Leu Cys Gly Val Val Gly Phe Arg Pro Thr Leu Ala Arg  
 180 185 190

Tyr Pro Arg Asp Arg Ile Ile Pro Val Ser Pro Thr Arg Asp Thr Ala  
 195 200 205

Gly Ile Ile Ala Gln Cys Val Ala Asp Val Ile Ile Leu Asp Gln Val  
 210 215 220

Ile Ser Gly Arg Ser Ala Lys Ile Ser Pro Met Pro Leu Lys Gly Leu  
 225 230 235 240

Arg Ile Gly Leu Pro Thr Thr Tyr Phe Tyr Asp Asp Leu Asp Ala Asp  
 245 250 255

Val Ala Phe Ala Ala Glu Thr Thr Ile Arg Leu Leu Ala Asn Arg Gly  
 260 265 270  
 Val Thr Phe Val Glu Ala Asp Ile Pro His Leu Glu Glu Leu Asn Ser  
 275 280 285  
 Gly Ala Ser Leu Pro Ile Ala Leu Tyr Glu Phe Pro His Ala Leu Lys  
 290 295 300  
 Lys Tyr Leu Asp Asp Phe Val Gly Thr Val Ser Phe Ser Asp Val Ile  
 305 310 315 320  
 Lys Gly Ile Arg Ser Pro Asp Val Ala Asn Ile Val Ser Ala Gln Ile  
 325 330 335  
 Asp Gly His Gln Ile Ser Asn Asp Glu Tyr Glu Leu Ala Arg Gln Ser  
 340 345 350  
 Phe Arg Pro Arg Leu Gln Ala Thr Tyr Arg Asn Tyr Phe Arg Leu Tyr  
 355 360 365  
 Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Ala Ala Lys Ala  
 370 375 380  
 Ile Gly Gln Glu Ser Ser Val Ile His Asn Gly Ser Met Met Asn Thr  
 385 390 395 400  
 Phe Lys Ile Tyr Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu  
 405 410 415  
 Pro Gly Leu Ser Leu Pro Ala Cys Leu Thr Pro Asp Arg Leu Pro Val  
 420 425 430  
 Gly Met Glu Ile Asp Gly Leu Ala Gly Ser Asp His Arg Leu Leu Ala  
 435 440 445  
 Ile Gly Ala Ala Leu Glu Lys Ala Ile Asn Phe Pro Ser Phe Pro Asp  
 450 455 460  
 Ala Phe Asn  
 465

&lt;210&gt; 9

&lt;211&gt; 1404

&lt;212&gt; DNA

&lt;213&gt; Agrobacterium tumefaciens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1401)

&lt;223&gt; coding for indole acetamide hydrolase (tms2)

&lt;400&gt; 9

atg gtg ccc att acc tcg tta gca caa acc cta gaa cgc ctg aga cgg 48  
 Met Val Pro Ile Thr Ser Leu Ala Gln Thr Leu Glu Arg Leu Arg Arg  
 1 5 10 15  
 aaa gac tac tcc tgc tta gaa cta gta gaa act ctg ata gcg cgt tgc 96  
 Lys Asp Tyr Ser Cys Leu Glu Leu Val Glu Thr Leu Ile Ala Arg Cys  
 20 25 30  
 caa gct gca aaa cca tta aat gcc ctt ctg gct aca gac tgg gat ggc 144  
 Gln Ala Ala Lys Pro Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Gly  
 35 40 45  
 ttg cgg cga agc gcc aaa aaa att gat cgt cat gga aac gcc gga tta 192  
 Leu Arg Arg Ser Ala Lys Lys Ile Asp Arg His Gly Asn Ala Gly Leu  
 50 55 60



ggt	ctt	tgc	ggc	att	cca	ctc	tgt	ttt	aag	gcg	aac	atc	gcg	acc	ggc	240
Gly	Leu	Cys	Gly	Ile	Pro	Leu	Cys	Phe	Lys	Ala	Asn	Ile	Ala	Thr	Gly	
65					70					75					80	
ata	ttt	cct	aca	agc	gct	gct	act	ccg	gcg	ctg	ata	aac	cac	ttg	cca	288
Ile	Phe	Pro	Thr	Ser	Ala	Ala	Thr	Pro	Ala	Leu	Ile	Asn	His	Leu	Pro	
				85				90						95		
aag	ata	cca	tcc	cgc	gtc	gca	gaa	aga	ctt	ttt	tca	gct	gga	gca	ctg	336
Lys	Ile	Pro	Ser	Arg	Val	Ala	Glu	Arg	Leu	Phe	Ser	Ala	Gly	Ala	Leu	
			100					105					110			
ccg	ggt	gcc	tcg	gga	aac	atg	cat	gag	tta	tcg	ttt	gga	att	acg	agc	384
Pro	Gly	Ala	Ser	Gly	Asn	Met	His	Glu	Leu	Ser	Phe	Gly	Ile	Thr	Ser	
		115					120					125				
aac	aac	tat	gcc	acc	ggt	gcg	gtg	cgg	aac	ccg	tgg	aat	cca	agt	ctg	432
Asn	Asn	Tyr	Ala	Thr	Gly	Ala	Val	Arg	Asn	Pro	Trp	Asn	Pro	Ser	Leu	
		130				135					140					
ata	cca	gga	ggc	tca	agc	ggt	ggt	gtg	gct	gct	gcg	gtg	gca	agc	cga	480
Ile	Pro	Gly	Gly	Ser	Ser	Gly	Gly	Val	Ala	Ala	Ala	Val	Ala	Ser	Arg	
145						150				155					160	
ttg	atg	tta	ggc	ggc	ata	ggc	acc	gat	acc	ggt	gca	tct	gtt	cgc	cta	528
Leu	Met	Leu	Gly	Gly	Ile	Gly	Thr	Asp	Thr	Gly	Ala	Ser	Val	Arg	Leu	
			165						170					175		
ccc	gca	gcc	ctg	tgt	ggc	gta	gta	gga	ttt	cga	ccg	acg	ctt	gct	cga	576
Pro	Ala	Ala	Leu	Cys	Gly	Val	Val	Gly	Phe	Arg	Pro	Thr	Leu	Ala	Arg	
			180					185					190			
tat	cca	aga	gat	cgg	ata	ata	ccg	gtc	agc	ccc	acc	cgg	gac	acc	gcc	624
Tyr	Pro	Arg	Asp	Arg	Ile	Ile	Pro	Val	Ser	Pro	Thr	Arg	Asp	Thr	Ala	
		195					200					205				
gga	atc	ata	gcg	cag	tgc	gta	gcc	gat	gtt	ata	atc	ctc	gat	cag	gtg	672
Gly	Ile	Ile	Ala	Gln	Cys	Val	Ala	Asp	Val	Ile	Ile	Leu	Asp	Gln	Val	
	210					215					220					
att	tcc	gga	cgg	tcg	gcg	aaa	att	tca	ccc	atg	ccg	ctg	aag	ggg	ctt	720
Ile	Ser	Gly	Arg	Ser	Ala	Lys	Ile	Ser	Pro	Met	Pro	Leu	Lys	Gly	Leu	
225					230					235					240	
cgg	atc	ggc	ctc	ccc	act	acc	tac	ttt	tac	gat	gac	ctt	gat	gct	gat	768
Arg	Ile	Gly	Leu	Pro	Thr	Thr	Tyr	Phe	Tyr	Asp	Asp	Leu	Asp	Ala	Asp	
				245					250					255		
gtg	gcc	ttc	gca	gct	gaa	acg	acg	att	cgc	ttg	cta	gcc	aac	aga	ggc	816
Val	Ala	Phe	Ala	Ala	Glu	Thr	Thr	Ile	Arg	Leu	Leu	Ala	Asn	Arg	Gly	
			260					265					270			
gta	acc	ttt	gtt	gaa	gcc	gac	atc	ccc	cac	cta	gag	gaa	ctg	aat	agt	864
Val	Thr	Phe	Val	Glu	Ala	Asp	Ile	Pro	His	Leu	Glu	Glu	Leu	Asn	Ser	
		275					280						285			
ggg	gca	agt	ttg	cca	att	gcg	ctt	tac	gaa	ttt	cca	cac	gct	cta	aaa	912
Gly	Ala	Ser	Leu	Pro	Ile	Ala	Leu	Tyr	Glu	Phe	Pro	His	Ala	Leu	Lys	
	290					295					300					
aag	tat	ctc	gac	gat	ttt	gtg	gga	aca	gtt	tct	ttt	tct	gac	gtt	atc	960
Lys	Tyr	Leu	Asp	Asp	Phe	Val	Gly	Thr	Val	Ser	Phe	Ser	Asp	Val	Ile	
305					310					315					320	
aaa	gga	att	cgt	agc	ccc	gat	gta	gcg	aac	att	gtc	agt	gcg	caa	att	1008
Lys	Gly	Ile	Arg	Ser	Pro	Asp	Val	Ala	Asn	Ile	Val	Ser	Ala	Gln	Ile	
				325					330					335		

gat ggg cat caa att tcc aac gat gaa tat gaa ctg gcg cgt caa tcc 1056  
Asp Gly His Gln Ile Ser Asn Asp Glu Tyr Glu Leu Ala Arg Gln Ser  
340 345 350  
ttc agg cca agg ctc cag gcc act tat cgg aat tac ttc aga ctc tat 1104  
Phe Arg Pro Arg Leu Gln Ala Thr Tyr Arg Asn Tyr Phe Arg Leu Tyr  
355 360 365  
cag tta gat gca atc ctt ttc cca act gca ccc tta gcg gcc aaa gcc 1152  
Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Ala Ala Lys Ala  
370 375 380  
ata ggt cag gag tcg tca gtc atc cac aat ggc tca atg ata aac act 1200  
Ile Gly Gln Glu Ser Ser Val Ile His Asn Gly Ser Met Ile Asn Thr  
385 390 395 400  
ttc aag atc tac gtg cga aat gtg gac cca agc agc aac gca ggc cta 1248  
Phe Lys Ile Tyr Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu  
405 410 415  
cct ggg ttg agc ctt cct gcc tgc ctt aca cct gat cgc ttg cct gtt 1296  
Pro Gly Leu Ser Leu Pro Ala Cys Leu Thr Pro Asp Arg Leu Pro Val  
420 425 430  
gga atg gaa att gac gga tta gcg ggg tca gac cac cgt ctg tta gca 1344  
Gly Met Glu Ile Asp Gly Leu Ala Gly Ser Asp His Arg Leu Leu Ala  
435 440 445  
atc ggg gca gca tta gaa aaa gcc ata aat ttt cct tcc ttt ccc gat 1392  
Ile Gly Ala Ala Leu Glu Lys Ala Ile Asn Phe Pro Ser Phe Pro Asp  
450 455 460  
gct ttt aat tag 1404  
Ala Phe Asn  
465  
<210> 10  
<211> 467  
<212> PRT  
<213> Agrobacterium tumefaciens  
<400> 10  
Met Val Pro Ile Thr Ser Leu Ala Gln Thr Leu Glu Arg Leu Arg Arg  
1 5 10 15  
Lys Asp Tyr Ser Cys Leu Glu Leu Val Glu Thr Leu Ile Ala Arg Cys  
20 25 30  
Gln Ala Ala Lys Pro Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Gly  
35 40 45  
Leu Arg Arg Ser Ala Lys Lys Ile Asp Arg His Gly Asn Ala Gly Leu  
50 55 60  
Gly Leu Cys Gly Ile Pro Leu Cys Phe Lys Ala Asn Ile Ala Thr Gly  
65 70 75 80  
Ile Phe Pro Thr Ser Ala Ala Thr Pro Ala Leu Ile Asn His Leu Pro  
85 90 95  
Lys Ile Pro Ser Arg Val Ala Glu Arg Leu Phe Ser Ala Gly Ala Leu  
100 105 110  
Pro Gly Ala Ser Gly Asn Met His Glu Leu Ser Phe Gly Ile Thr Ser  
115 120 125  
Asn Asn Tyr Ala Thr Gly Ala Val Arg Asn Pro Trp Asn Pro Ser Leu  
130 135 140

Ile Pro Gly Gly Ser Ser Gly Gly Val Ala Ala Ala Val Ala Ser Arg  
 145 150 155 160  
 Leu Met Leu Gly Gly Ile Gly Thr Asp Thr Gly Ala Ser Val Arg Leu  
 165 170 175  
 Pro Ala Ala Leu Cys Gly Val Val Gly Phe Arg Pro Thr Leu Ala Arg  
 180 185 190  
 Tyr Pro Arg Asp Arg Ile Ile Pro Val Ser Pro Thr Arg Asp Thr Ala  
 195 200 205  
 Gly Ile Ile Ala Gln Cys Val Ala Asp Val Ile Ile Leu Asp Gln Val  
 210 215 220  
 Ile Ser Gly Arg Ser Ala Lys Ile Ser Pro Met Pro Leu Lys Gly Leu  
 225 230 235 240  
 Arg Ile Gly Leu Pro Thr Thr Tyr Phe Tyr Asp Asp Leu Asp Ala Asp  
 245 250 255  
 Val Ala Phe Ala Ala Glu Thr Thr Ile Arg Leu Leu Ala Asn Arg Gly  
 260 265 270  
 Val Thr Phe Val Glu Ala Asp Ile Pro His Leu Glu Glu Leu Asn Ser  
 275 280 285  
 Gly Ala Ser Leu Pro Ile Ala Leu Tyr Glu Phe Pro His Ala Leu Lys  
 290 295 300  
 Lys Tyr Leu Asp Asp Phe Val Gly Thr Val Ser Phe Ser Asp Val Ile  
 305 310 315 320  
 Lys Gly Ile Arg Ser Pro Asp Val Ala Asn Ile Val Ser Ala Gln Ile  
 325 330 335  
 Asp Gly His Gln Ile Ser Asn Asp Glu Tyr Glu Leu Ala Arg Gln Ser  
 340 345 350  
 Phe Arg Pro Arg Leu Gln Ala Thr Tyr Arg Asn Tyr Phe Arg Leu Tyr  
 355 360 365  
 Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Ala Ala Lys Ala  
 370 375 380  
 Ile Gly Gln Glu Ser Ser Val Ile His Asn Gly Ser Met Ile Asn Thr  
 385 390 395 400  
 Phe Lys Ile Tyr Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu  
 405 410 415  
 Pro Gly Leu Ser Leu Pro Ala Cys Leu Thr Pro Asp Arg Leu Pro Val  
 420 425 430  
 Gly Met Glu Ile Asp Gly Leu Ala Gly Ser Asp His Arg Leu Leu Ala  
 435 440 445  
 Ile Gly Ala Ala Leu Glu Lys Ala Ile Asn Phe Pro Ser Phe Pro Asp  
 450 455 460  
 Ala Phe Asn  
 465

&lt;210&gt; 11

&lt;211&gt; 609

&lt;212&gt; DNA

&lt;213&gt; Xanthobacter autotrophicus

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(603)

&lt;223&gt; coding for haloalkane dehalohenase

&lt;400&gt; 11

```

atg tca acg ttt ttt gaa ccg gag aac gga atg aaa caa aac gcc aaa 48
Met Ser Thr Phe Phe Glu Pro Glu Asn Gly Met Lys Gln Asn Ala Lys
  1             5             10             15

acc gaa cga atc ctg gat gtc gcg ctc gaa ttg ctt gag aca gag ggt 96
Thr Glu Arg Ile Leu Asp Val Ala Leu Glu Leu Leu Glu Thr Glu Gly
             20             25             30

gag ttt ggt ttg acg atg agg cag gtg gca acg caa gcg gac atg tcc 144
Glu Phe Gly Leu Thr Met Arg Gln Val Ala Thr Gln Ala Asp Met Ser
             35             40             45

ctg agc aac gtt cag tac tat ttc aag tcc gag gac ctg ctc ctc gtg 192
Leu Ser Asn Val Gln Tyr Tyr Phe Lys Ser Glu Asp Leu Leu Leu Val
             50             55             60

gcc atg gca gac cgt tac ttt caa cgg tgc ctg aca acc atg gct gag 240
Ala Met Ala Asp Arg Tyr Phe Gln Arg Cys Leu Thr Thr Met Ala Glu
             65             70             75             80

cat ccg ccc tta tcg gca ggg cgt gat caa cac gcc cag tta aga gcg 288
His Pro Pro Leu Ser Ala Gly Arg Asp Gln His Ala Gln Leu Arg Ala
             85             90             95

ttg tta cga gaa ctg ctc ggt cat ggt ctt gag att tcc gag atg tgt 336
Leu Leu Arg Glu Leu Leu Gly His Gly Leu Glu Ile Ser Glu Met Cys
             100             105             110

cga ata ttc agg gag tac tgg gca atc gcc acc cgt aat gaa act gtt 384
Arg Ile Phe Arg Glu Tyr Trp Ala Ile Ala Thr Arg Asn Glu Thr Val
             115             120             125

cac ggc tat ctc aag tcg tac tat cgg gat ctc gcc gaa gtg atg gct 432
His Gly Tyr Leu Lys Ser Tyr Tyr Arg Asp Leu Ala Glu Val Met Ala
             130             135             140

gag aag ctt gcg cca ctg gcc agc agc gaa aag gcg ctg gcc gtg gcc 480
Glu Lys Leu Ala Pro Leu Ala Ser Ser Glu Lys Ala Leu Ala Val Ala
             145             150             155             160

gta tct ttg gtt att cct tat gtt gag ggg tat tcg gta acg gcc att 528
Val Ser Leu Val Ile Pro Tyr Val Glu Gly Tyr Ser Val Thr Ala Ile
             165             170             175

gca atg ccc gaa tcc att gat acg att tcc gag acg ctg acc aat gtg 576
Ala Met Pro Glu Ser Ile Asp Thr Ile Ser Glu Thr Leu Thr Asn Val
             180             185             190

gtg ttg gag cag ctt cgc atc agc aat tcatga 609
Val Leu Glu Gln Leu Arg Ile Ser Asn
             195             200

```

&lt;210&gt; 12

&lt;211&gt; 201

&lt;212&gt; PRT

&lt;213&gt; Xanthobacter autotrophicus

&lt;400&gt; 12

```

Met Ser Thr Phe Phe Glu Pro Glu Asn Gly Met Lys Gln Asn Ala Lys
  1             5             10             15

```

Thr Glu Arg Ile Leu Asp Val Ala Leu Glu Leu Leu Glu Thr Glu Gly  
                   20                  25                  30  
 Glu Phe Gly Leu Thr Met Arg Gln Val Ala Thr Gln Ala Asp Met Ser  
           35                  40                  45  
 Leu Ser Asn Val Gln Tyr Tyr Phe Lys Ser Glu Asp Leu Leu Leu Val  
           50                  55                  60  
 Ala Met Ala Asp Arg Tyr Phe Gln Arg Cys Leu Thr Thr Met Ala Glu  
           65                  70                  75                  80  
 His Pro Pro Leu Ser Ala Gly Arg Asp Gln His Ala Gln Leu Arg Ala  
                   85                  90                  95  
 Leu Leu Arg Glu Leu Leu Gly His Gly Leu Glu Ile Ser Glu Met Cys  
                   100                  105                  110  
 Arg Ile Phe Arg Glu Tyr Trp Ala Ile Ala Thr Arg Asn Glu Thr Val  
           115                  120                  125  
 His Gly Tyr Leu Lys Ser Tyr Tyr Arg Asp Leu Ala Glu Val Met Ala  
           130                  135                  140  
 Glu Lys Leu Ala Pro Leu Ala Ser Ser Glu Lys Ala Leu Ala Val Ala  
           145                  150                  155                  160  
 Val Ser Leu Val Ile Pro Tyr Val Glu Gly Tyr Ser Val Thr Ala Ile  
                   165                  170                  175  
 Ala Met Pro Glu Ser Ile Asp Thr Ile Ser Glu Thr Leu Thr Asn Val  
           180                  185                  190  
 Val Leu Glu Gln Leu Arg Ile Ser Asn  
           195                  200

&lt;210&gt; 13

&lt;211&gt; 1131

&lt;212&gt; DNA

&lt;213&gt; Herpes simplex virus 1

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1128)

&lt;223&gt; coding for thymidine kinase (TK)

&lt;400&gt; 13

atg gct tcg tac ccc tgc cat caa cac gcg tct gcg ttc gac cag gct 48  
 Met Ala Ser Tyr Pro Cys His Gln His Ala Ser Ala Phe Asp Gln Ala  
   1                  5                  10                  15  
 gcg cgt tct cgc ggc cat agc aac cga cgt acg gcg ttg cgc cct cgc 96  
 Ala Arg Ser Arg Gly His Ser Asn Arg Arg Thr Ala Leu Arg Pro Arg  
           20                  25                  30  
 cgg cag caa gaa gcc acg gaa gtc cgc ctg gag cag aaa atg ccc acg 144  
 Arg Gln Gln Glu Ala Thr Glu Val Arg Leu Glu Gln Lys Met Pro Thr  
           35                  40                  45  
 cta ctg cgg gtt tat ata gac ggt cct cac ggg atg ggg aaa acc acc 192  
 Leu Leu Arg Val Tyr Ile Asp Gly Pro His Gly Met Gly Lys Thr Thr  
           50                  55                  60  
 acc acg caa ctg ctg gtg gcc ctg ggt tcg cgc gac gat atc gtc tac 240  
 Thr Thr Gln Leu Leu Val Ala Leu Gly Ser Arg Asp Asp Ile Val Tyr  
           65                  70                  75                  80

gta ccc gag ccg atg act tac tgg cag gtg ctg ggg gct tcc gag aca	288
Val Pro Glu Pro Met Thr Tyr Trp Gln Val Leu Gly Ala Ser Glu Thr	
85 90 95	
atc gcg aac atc tac acc aca caa cac cgc ctc gac cag ggt gag ata	336
Ile Ala Asn Ile Tyr Thr Thr Gln His Arg Leu Asp Gln Gly Glu Ile	
100 105 110	
tcg gcc ggg gac gcg gcg gtg gta atg aca agc gcc cag ata aca atg	384
Ser Ala Gly Asp Ala Ala Val Val Met Thr Ser Ala Gln Ile Thr Met	
115 120 125	
ggc atg cct tat gcc gtg acc gac gcc gtt ctg gct cct cat gtc ggg	432
Gly Met Pro Tyr Ala Val Thr Asp Ala Val Leu Ala Pro His Val Gly	
130 135 140	
ggg gag gct ggg agt tca cat gcc ccg ccc ccg gcc ctc acc ctc atc	480
Gly Glu Ala Gly Ser Ser His Ala Pro Pro Pro Ala Leu Thr Leu Ile	
145 150 155 160	
ttc gac cgc cat ccc atc gcc gcc ctc ctg tgc tac ccg gcc gcg cga	528
Phe Asp Arg His Pro Ile Ala Ala Leu Leu Cys Tyr Pro Ala Ala Arg	
165 170 175	
tac ctt atg ggc agc atg acc ccc cag gcc gtg ctg gcg ttc gtg gcc	576
Tyr Leu Met Gly Ser Met Thr Pro Gln Ala Val Leu Ala Phe Val Ala	
180 185 190	
ctc atc ccg ccg acc ttg ccc ggc aca aac atc gtg ttg ggg gcc ctt	624
Leu Ile Pro Pro Thr Leu Pro Gly Thr Asn Ile Val Leu Gly Ala Leu	
195 200 205	
ccg gag gac aga cac atc gac cgc ctg gcc aaa cgc cag cgc ccc ggc	672
Pro Glu Asp Arg His Ile Asp Arg Leu Ala Lys Arg Gln Arg Pro Gly	
210 215 220	
gag cgg ctt gac ctg gct atg ctg gcc gcg att cgc cgc gtt tac ggg	720
Glu Arg Leu Asp Leu Ala Met Leu Ala Ala Ile Arg Arg Val Tyr Gly	
225 230 235 240	
ctg ctt gcc aat acg gtg cgg tat ctg cag ggc ggc ggg tcg tgg tgg	768
Leu Leu Ala Asn Thr Val Arg Tyr Leu Gln Gly Gly Gly Ser Trp Trp	
245 250 255	
gag gat tgg gga cag ctt tcg ggg acg gcc gtg ccg ccc cag ggt gcc	816
Glu Asp Trp Gly Gln Leu Ser Gly Thr Ala Val Pro Pro Gln Gly Ala	
260 265 270	
gag ccc cag agc aac gcg ggc cca cga ccc cat atc ggg gac acg tta	864
Glu Pro Gln Ser Asn Ala Gly Pro Arg Pro His Ile Gly Asp Thr Leu	
275 280 285	
ttt acc ctg ttt cgg gcc ccc gag ttg ctg gcc ccc aac ggc gac ctg	912
Phe Thr Leu Phe Arg Ala Pro Glu Leu Leu Ala Pro Asn Gly Asp Leu	
290 295 300	
tat aac gtg ttt gcc tgg gcc ttg gac gtc ttg gcc aaa cgc ctc cgt	960
Tyr Asn Val Phe Ala Trp Ala Leu Asp Val Leu Ala Lys Arg Leu Arg	
305 310 315 320	
ccc atg cac gtc ttt atc ctg gat tac gac caa tcg ccc gcc ggc tgc	1008
Pro Met His Val Phe Ile Leu Asp Tyr Asp Gln Ser Pro Ala Gly Cys	
325 330 335	
cgg gac gcc ctg ctg caa ctt acc tcc ggg atg gtc cag acc cac gtc	1056
Arg Asp Ala Leu Leu Gln Leu Thr Ser Gly Met Val Gln Thr His Val	
340 345 350	

acc acc cca ggc tcc ata ccg acg atc tgc gac ctg gcg cgc acg ttt 1104  
 Thr Thr Pro Gly Ser Ile Pro Thr Ile Cys Asp Leu Ala Arg Thr Phe  
 355 360 365

gcc cgg gag atg ggg gag gct aac tga 1131  
 Ala Arg Glu Met Gly Glu Ala Asn  
 370 375

<210> 14

<211> 376

<212> PRT

<213> Herpes simplex virus 1

<400> 14

Met Ala Ser Tyr Pro Cys His Gln His Ala Ser Ala Phe Asp Gln Ala  
 1 5 10 15  
 Ala Arg Ser Arg Gly His Ser Asn Arg Arg Thr Ala Leu Arg Pro Arg  
 20 25 30  
 Arg Gln Gln Glu Ala Thr Glu Val Arg Leu Glu Gln Lys Met Pro Thr  
 35 40 45  
 Leu Leu Arg Val Tyr Ile Asp Gly Pro His Gly Met Gly Lys Thr Thr  
 50 55 60  
 Thr Thr Gln Leu Leu Val Ala Leu Gly Ser Arg Asp Asp Ile Val Tyr  
 65 70 75 80  
 Val Pro Glu Pro Met Thr Tyr Trp Gln Val Leu Gly Ala Ser Glu Thr  
 85 90 95  
 Ile Ala Asn Ile Tyr Thr Thr Gln His Arg Leu Asp Gln Gly Glu Ile  
 100 105 110  
 Ser Ala Gly Asp Ala Ala Val Val Met Thr Ser Ala Gln Ile Thr Met  
 115 120 125  
 Gly Met Pro Tyr Ala Val Thr Asp Ala Val Leu Ala Pro His Val Gly  
 130 135 140  
 Gly Glu Ala Gly Ser Ser His Ala Pro Pro Pro Ala Leu Thr Leu Ile  
 145 150 155 160  
 Phe Asp Arg His Pro Ile Ala Ala Leu Leu Cys Tyr Pro Ala Ala Arg  
 165 170 175  
 Tyr Leu Met Gly Ser Met Thr Pro Gln Ala Val Leu Ala Phe Val Ala  
 180 185 190  
 Leu Ile Pro Pro Thr Leu Pro Gly Thr Asn Ile Val Leu Gly Ala Leu  
 195 200 205  
 Pro Glu Asp Arg His Ile Asp Arg Leu Ala Lys Arg Gln Arg Pro Gly  
 210 215 220  
 Glu Arg Leu Asp Leu Ala Met Leu Ala Ala Ile Arg Arg Val Tyr Gly  
 225 230 235 240  
 Leu Leu Ala Asn Thr Val Arg Tyr Leu Gln Gly Gly Gly Ser Trp Trp  
 245 250 255  
 Glu Asp Trp Gly Gln Leu Ser Gly Thr Ala Val Pro Pro Gln Gly Ala  
 260 265 270  
 Glu Pro Gln Ser Asn Ala Gly Pro Arg Pro His Ile Gly Asp Thr Leu  
 275 280 285  
 Phe Thr Leu Phe Arg Ala Pro Glu Leu Leu Ala Pro Asn Gly Asp Leu  
 290 295 300

Tyr Asn Val Phe Ala Trp Ala Leu Asp Val Leu Ala Lys Arg Leu Arg  
 305 310 315 320  
 Pro Met His Val Phe Ile Leu Asp Tyr Asp Gln Ser Pro Ala Gly Cys  
 325 330 335  
 Arg Asp Ala Leu Leu Gln Leu Thr Ser Gly Met Val Gln Thr His Val  
 340 345 350  
 Thr Thr Pro Gly Ser Ile Pro Thr Ile Cys Asp Leu Ala Arg Thr Phe  
 355 360 365  
 Ala Arg Glu Met Gly Glu Ala Asn  
 370 375

&lt;210&gt; 15

&lt;211&gt; 1131

&lt;212&gt; DNA

&lt;213&gt; Herpes simplex virus 1

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1128)

&lt;223&gt; coding for thymidine kinase (TK)

&lt;400&gt; 15

atg gct tcg tac ccc tgc cat caa cac gcg tct gcg ttc gac cag gct	48
Met Ala Ser Tyr Pro Cys His Gln His Ala Ser Ala Phe Asp Gln Ala	
1 5 10 15	
gcg cgt tct cgc ggc cat agc aac cga cgt acg gcg ttg cgc cct cgc	96
Ala Arg Ser Arg Gly His Ser Asn Arg Arg Thr Ala Leu Arg Pro Arg	
20 25 30	
cgg cag caa gaa gcc acg gaa gtc cgc ctg gag cag aaa atg ccc acg	144
Arg Gln Gln Glu Ala Thr Glu Val Arg Leu Glu Gln Lys Met Pro Thr	
35 40 45	
cta ctg cgg gtt tat ata gac ggt cct cac ggg atg ggg aaa acc acc	192
Leu Leu Arg Val Tyr Ile Asp Gly Pro His Gly Met Gly Lys Thr Thr	
50 55 60	
acc acg caa ctg ctg gtg gcc ctg ggt tcg cgc gac gat atc gtc tac	240
Thr Thr Gln Leu Leu Val Ala Leu Gly Ser Arg Asp Asp Ile Val Tyr	
65 70 75 80	
gta ccc gag ccg atg act tac tgg cag gtg ctg ggg gct tcc gag aca	288
Val Pro Glu Pro Met Thr Tyr Trp Gln Val Leu Gly Ala Ser Glu Thr	
85 90 95	
atc gcg aac atc tac acc aca caa cac cgc ctc gac cag ggt gag ata	336
Ile Ala Asn Ile Tyr Thr Thr Gln His Arg Leu Asp Gln Gly Glu Ile	
100 105 110	
tcg gcc ggg gac gcg gcg gtg gta atg aca agc gcc cag ata aca atg	384
Ser Ala Gly Asp Ala Ala Val Val Met Thr Ser Ala Gln Ile Thr Met	
115 120 125	
ggc atg cct tat gcc gtg acc gac gcc gtt ctg gct cct cat gtc ggg	432
Gly Met Pro Tyr Ala Val Thr Asp Ala Val Leu Ala Pro His Val Gly	
130 135 140	
ggg gag gct ggg agt tca cat gcc ccg ccc ccg gcc ctc acc ctc atc	480
Gly Glu Ala Gly Ser Ser His Ala Pro Pro Pro Ala Leu Thr Leu Ile	
145 150 155 160	



```

ttc gac cgc cat ccc atc gcc gcc ctc ctg tgc tac ccg gcc gcg cga 528
Phe Asp Arg His Pro Ile Ala Ala Leu Leu Cys Tyr Pro Ala Ala Arg
165 170 175

tac ctt atg ggc agc atg acc ccc cag gcc gtg ctg gcg ttc gtg gcc 576
Tyr Leu Met Gly Ser Met Thr Pro Gln Ala Val Leu Ala Phe Val Ala
180 185 190

ctc atc ccg ccg acc ttg ccc ggc aca aac atc gtg ttg ggg gcc ctt 624
Leu Ile Pro Pro Thr Leu Pro Gly Thr Asn Ile Val Leu Gly Ala Leu
195 200 205

ccg gag gac aga cac atc gac cgc ctg gcc aaa cgc cag cgc ccc ggc 672
Pro Glu Asp Arg His Ile Asp Arg Leu Ala Lys Arg Gln Arg Pro Gly
210 215 220

gag cgg ctt gac ctg gct atg ctg gcc gcg att cgc cgc gtt tac ggg 720
Glu Arg Leu Asp Leu Ala Met Leu Ala Ala Ile Arg Arg Val Tyr Gly
225 230 235

ctg ctt gcc aat acg gtg cgg tat ctg cag ggc ggc ggg tcg tgg tgg 768
Leu Leu Ala Asn Thr Val Arg Tyr Leu Gln Gly Gly Gly Ser Trp Trp
245 250 255

gag gat tgg gga cag ctt tcg ggg acg gcc gtg ccg ccc cag ggt gcc 816
Glu Asp Trp Gly Gln Leu Ser Gly Thr Ala Val Pro Pro Gln Gly Ala
260 265 270

gag ccc cag agc aac gcg ggc cca cga ccc cat atc ggg gac acg tta 864
Glu Pro Gln Ser Asn Ala Gly Pro Arg Pro His Ile Gly Asp Thr Leu
275 280 285

ttt acc ctg ttt cgg gcc ccc gag ttg ctg gcc ccc aac ggc gac ctg 912
Phe Thr Leu Phe Arg Ala Pro Glu Leu Leu Ala Pro Asn Gly Asp Leu
290 295 300

tat aac gtg ttt gcc tgg gcc ttg gac gtc ttg gcc aaa cgc ctc cgt 960
Tyr Asn Val Phe Ala Trp Ala Leu Asp Val Leu Ala Lys Arg Leu Arg
305 310 315

ccc atg cac gtc ttt atc ctg gat tac gac caa tcg ccc gcc ggc tgc 1008
Pro Met His Val Phe Ile Leu Asp Tyr Asp Gln Ser Pro Ala Gly Cys
325 330 335

cgg gac gcc ctg ctg caa ctt acc tcc ggg atg gtc cag acc cac gtc 1056
Arg Asp Ala Leu Leu Gln Leu Thr Ser Gly Met Val Gln Thr His Val
340 345 350

acc acc cca ggc tcc ata ccg acg atc tgc gac ctg gcg cgc acg ttt 1104
Thr Thr Pro Gly Ser Ile Pro Thr Ile Cys Asp Leu Ala Arg Thr Phe
355 360 365

gcc cgg gag atg ggg gag gct aac tga 1131
Ala Arg Glu Met Gly Glu Ala Asn
370 375

<210> 16
<211> 376
<212> PRT
<213> Herpes simplex virus 1

<400> 16
Met Ala Ser Tyr Pro Cys His Gln His Ala Ser Ala Phe Asp Gln Ala
1 5 10 15
Ala Arg Ser Arg Gly His Ser Asn Arg Arg Thr Ala Leu Arg Pro Arg
20 25 30

```

Arg Gln Gln Glu Ala Thr Glu Val Arg Leu Glu Gln Lys Met Pro Thr  
                   35                                  40                                  45  
 Leu Leu Arg Val Tyr Ile Asp Gly Pro His Gly Met Gly Lys Thr Thr  
           50                                  55                                  60  
 Thr Thr Gln Leu Leu Val Ala Leu Gly Ser Arg Asp Asp Ile Val Tyr  
   65                                  70                                  75                                  80  
 Val Pro Glu Pro Met Thr Tyr Trp Gln Val Leu Gly Ala Ser Glu Thr  
                                   85                                  90                                  95  
 Ile Ala Asn Ile Tyr Thr Thr Gln His Arg Leu Asp Gln Gly Glu Ile  
                                  100                                 105                                 110  
 Ser Ala Gly Asp Ala Ala Val Val Met Thr Ser Ala Gln Ile Thr Met  
          115                                 120                                 125  
 Gly Met Pro Tyr Ala Val Thr Asp Ala Val Leu Ala Pro His Val Gly  
   130                                 135                                 140  
 Gly Glu Ala Gly Ser Ser His Ala Pro Pro Pro Ala Leu Thr Leu Ile  
  145                                 150                                 155                                 160  
 Phe Asp Arg His Pro Ile Ala Ala Leu Leu Cys Tyr Pro Ala Ala Arg  
                                  165                                 170                                 175  
 Tyr Leu Met Gly Ser Met Thr Pro Gln Ala Val Leu Ala Phe Val Ala  
          180                                 185                                 190  
 Leu Ile Pro Pro Thr Leu Pro Gly Thr Asn Ile Val Leu Gly Ala Leu  
          195                                 200                                 205  
 Pro Glu Asp Arg His Ile Asp Arg Leu Ala Lys Arg Gln Arg Pro Gly  
   210                                 215                                 220  
 Glu Arg Leu Asp Leu Ala Met Leu Ala Ala Ile Arg Arg Val Tyr Gly  
  225                                 230                                 235                                 240  
 Leu Leu Ala Asn Thr Val Arg Tyr Leu Gln Gly Gly Gly Ser Trp Trp  
                                  245                                 250                                 255  
 Glu Asp Trp Gly Gln Leu Ser Gly Thr Ala Val Pro Pro Gln Gly Ala  
          260                                 265                                 270  
 Glu Pro Gln Ser Asn Ala Gly Pro Arg Pro His Ile Gly Asp Thr Leu  
          275                                 280                                 285  
 Phe Thr Leu Phe Arg Ala Pro Glu Leu Leu Ala Pro Asn Gly Asp Leu  
   290                                 295                                 300  
 Tyr Asn Val Phe Ala Trp Ala Leu Asp Val Leu Ala Lys Arg Leu Arg  
  305                                 310                                 315                                 320  
 Pro Met His Val Phe Ile Leu Asp Tyr Asp Gln Ser Pro Ala Gly Cys  
                                  325                                 330                                 335  
 Arg Asp Ala Leu Leu Gln Leu Thr Ser Gly Met Val Gln Thr His Val  
          340                                 345                                 350  
 Thr Thr Pro Gly Ser Ile Pro Thr Ile Cys Asp Leu Ala Arg Thr Phe  
          355                                 360                                 365  
 Ala Arg Glu Met Gly Glu Ala Asn  
   370                                 375

&lt;210&gt; 17

&lt;211&gt; 840

&lt;212&gt; DNA

<213> *Toxoplasma gondii*

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(837)

<223> coding for hypoxanthine-xanthine-guanine  
phosphoribosyl transferase (HXGPRTase)

&lt;400&gt; 17

atg gcg tcc aaa ccc att gaa gaa tcc cgg tcg caa aaa cgg agt gcc	48
Met Ala Ser Lys Pro Ile Glu Glu Ser Arg Ser Gln Lys Arg Ser Ala	
1 5 10 15	
ttc tca gac atc ttc tgt tgt tgc act cct aat gaa ggg gct atc gtg	96
Phe Ser Asp Ile Phe Cys Cys Cys Thr Pro Asn Glu Gly Ala Ile Val	
20 25 30	
ccc agt gac cca atg gtc tcc acc agt gct cca gca cgc acc agt gct	144
Pro Ser Asp Pro Met Val Ser Thr Ser Ala Pro Ala Arg Thr Ser Ala	
35 40 45	
cca gcg cgc tcc agt gca ctt caa gac tac ggc aag ggc aag ggc cgt	192
Pro Ala Arg Ser Ser Ala Leu Gln Asp Tyr Gly Lys Gly Lys Gly Arg	
50 55 60	
att gag ccc atg tat atc ccc gac aac acc ttc tac aac gct gat gac	240
Ile Glu Pro Met Tyr Ile Pro Asp Asn Thr Phe Tyr Asn Ala Asp Asp	
65 70 75 80	
ttt ctt gtg ccc ccc cac tgc aag ccc tac att gac aaa atc ctc ctc	288
Phe Leu Val Pro Pro His Cys Lys Pro Tyr Ile Asp Lys Ile Leu Leu	
85 90 95	
cct ggt gga ttg gtc aag gac aga gtt gag aag ttg gcg tat gac atc	336
Pro Gly Gly Leu Val Lys Asp Arg Val Glu Lys Leu Ala Tyr Asp Ile	
100 105 110	
cac aga act tac ttc ggc gag gag ttg cac atc att tgc atc ctg aaa	384
His Arg Thr Tyr Phe Gly Glu Glu Leu His Ile Ile Cys Ile Leu Lys	
115 120 125	
ggc tct cgc ggc ttc ttc aac ctt ctg atc gac tac ctt gcc acc ata	432
Gly Ser Arg Gly Phe Phe Asn Leu Leu Ile Asp Tyr Leu Ala Thr Ile	
130 135 140	
cag aag tac agt ggt cgt gag tcc agc gtg ccc ccc ttc ttc gag cac	480
Gln Lys Tyr Ser Gly Arg Glu Ser Ser Val Pro Pro Phe Phe Glu His	
145 150 155 160	
tat gtc cgc ctg aag tcc tac cag aac gac aac agc aca ggc cag ctc	528
Tyr Val Arg Leu Lys Ser Tyr Gln Asn Asp Asn Ser Thr Gly Gln Leu	
165 170 175	
acc gtc ttg agc gac gac ttg tca atc ttt cgc gac aag cac gtt ctg	576
Thr Val Leu Ser Asp Asp Leu Ser Ile Phe Arg Asp Lys His Val Leu	
180 185 190	
att gtt gag gac atc gtc gac acc ggt ttc acc ctc acc gag ttc ggt	624
Ile Val Glu Asp Ile Val Asp Thr Gly Phe Thr Leu Thr Glu Phe Gly	
195 200 205	
gag cgc ctg aaa gcc gtc ggt ccc aag tcg atg aga atc gcc acc ctc	672
Glu Arg Leu Lys Ala Val Gly Pro Lys Ser Met Arg Ile Ala Thr Leu	
210 215 220	
gtc gag aag cgc aca gat cgc tcc aac agc ttg aag ggc gac ttc gtc	720
Val Glu Lys Arg Thr Asp Arg Ser Asn Ser Leu Lys Gly Asp Phe Val	
225 230 235 240	

ggc ttc agc att gaa gac gtc tgg atc gtt ggt tgc tgc tac gac ttc 768  
 Gly Phe Ser Ile Glu Asp Val Trp Ile Val Gly Cys Cys Tyr Asp Phe  
                   245                  250                  255

aac gag atg ttc cgc gac ttc gac cac gtc gcc gtc ctg agc gac gcc 816  
 Asn Glu Met Phe Arg Asp Phe Asp His Val Ala Val Leu Ser Asp Ala  
                   260                  265                  270

gct cgc aaa aag ttc gag aag taa 840  
 Ala Arg Lys Lys Phe Glu Lys  
                   275

<210> 18

<211> 279

<212> PRT

<213> Toxoplasma gondii

<400> 18

Met Ala Ser Lys Pro Ile Glu Glu Ser Arg Ser Gln Lys Arg Ser Ala  
   1                  5                  10                  15

Phe Ser Asp Ile Phe Cys Cys Cys Thr Pro Asn Glu Gly Ala Ile Val  
                   20                  25                  30

Pro Ser Asp Pro Met Val Ser Thr Ser Ala Pro Ala Arg Thr Ser Ala  
                   35                  40                  45

Pro Ala Arg Ser Ser Ala Leu Gln Asp Tyr Gly Lys Gly Lys Gly Arg  
                   50                  55                  60

Ile Glu Pro Met Tyr Ile Pro Asp Asn Thr Phe Tyr Asn Ala Asp Asp  
                   65                  70                  75                  80

Phe Leu Val Pro Pro His Cys Lys Pro Tyr Ile Asp Lys Ile Leu Leu  
                   85                  90                  95

Pro Gly Gly Leu Val Lys Asp Arg Val Glu Lys Leu Ala Tyr Asp Ile  
                   100                  105                  110

His Arg Thr Tyr Phe Gly Glu Glu Leu His Ile Ile Cys Ile Leu Lys  
                   115                  120                  125

Gly Ser Arg Gly Phe Phe Asn Leu Leu Ile Asp Tyr Leu Ala Thr Ile  
                   130                  135                  140

Gln Lys Tyr Ser Gly Arg Glu Ser Ser Val Pro Pro Phe Phe Glu His  
                   145                  150                  155                  160

Tyr Val Arg Leu Lys Ser Tyr Gln Asn Asp Asn Ser Thr Gly Gln Leu  
                   165                  170                  175

Thr Val Leu Ser Asp Asp Leu Ser Ile Phe Arg Asp Lys His Val Leu  
                   180                  185                  190

Ile Val Glu Asp Ile Val Asp Thr Gly Phe Thr Leu Thr Glu Phe Gly  
                   195                  200                  205

Glu Arg Leu Lys Ala Val Gly Pro Lys Ser Met Arg Ile Ala Thr Leu  
                   210                  215                  220

Val Glu Lys Arg Thr Asp Arg Ser Asn Ser Leu Lys Gly Asp Phe Val  
                   225                  230                  235                  240

Gly Phe Ser Ile Glu Asp Val Trp Ile Val Gly Cys Cys Tyr Asp Phe  
                   245                  250                  255

Asn Glu Met Phe Arg Asp Phe Asp His Val Ala Val Leu Ser Asp Ala  
                   260                  265                  270

Ala Arg Lys Lys Phe Glu Lys  
275

<210> 19

<211> 459

<212> DNA

<213> Escherichia coli

<220>

<221> CDS

<222> (1)..(456)

<223> coding for xanthine-guanine phosphoribosyl  
transferase (gpt)

<400> 19

atg agc gaa aaa tac atc gtc acc tgg gac atg ttg cag atc cat gca	48
Met Ser Glu Lys Tyr Ile Val Thr Trp Asp Met Leu Gln Ile His Ala	
1 5 10 15	
cgt aaa ctc gca agc cga ctg atg cct tct gaa caa tgg aaa ggc att	96
Arg Lys Leu Ala Ser Arg Leu Met Pro Ser Glu Gln Trp Lys Gly Ile	
20 25 30	
att gcc gta agc cgt ggc ggt ctg gta ccg ggt gcg tta ctg gcg cgt	144
Ile Ala Val Ser Arg Gly Gly Leu Val Pro Gly Ala Leu Leu Ala Arg	
35 40 45	
gaa ctg ggt att cgt cat gtc gat acc gtt tgt att tcc agc tac gat	192
Glu Leu Gly Ile Arg His Val Asp Thr Val Cys Ile Ser Ser Tyr Asp	
50 55 60	
cac gac aac cag cgc gag ctt aaa gtg ctg aaa cgc gca gaa ggc gat	240
His Asp Asn Gln Arg Glu Leu Lys Val Leu Lys Arg Ala Glu Gly Asp	
65 70 75 80	
ggc gaa ggc ttc atc gtt att gat gac ctg gtg gat acc ggt ggt act	288
Gly Glu Gly Phe Ile Val Ile Asp Asp Leu Val Asp Thr Gly Gly Thr	
85 90 95	
gcg gtt gcg att cgt gaa atg tat cca aaa gcg cac ttt gtc acc atc	336
Ala Val Ala Ile Arg Glu Met Tyr Pro Lys Ala His Phe Val Thr Ile	
100 105 110	
ttc gca aaa ccg gct ggt cgt ccg ctg gtt gat gac tat gtt gtt gat	384
Phe Ala Lys Pro Ala Gly Arg Pro Leu Val Asp Asp Tyr Val Val Asp	
115 120 125	
atc ccg caa gat acc tgg att gaa cag ccg tgg gat atg ggc gtc gta	432
Ile Pro Gln Asp Thr Trp Ile Glu Gln Pro Trp Asp Met Gly Val Val	
130 135 140	
ttc gtc ccg cca atc tcc ggt cgc taa	459
Phe Val Pro Pro Ile Ser Gly Arg	
145 150	

<210> 20

<211> 152

<212> PRT

<213> Escherichia coli

<400> 20

Met Ser Glu Lys Tyr Ile Val Thr Trp Asp Met Leu Gln Ile His Ala	
1 5 10 15	
Arg Lys Leu Ala Ser Arg Leu Met Pro Ser Glu Gln Trp Lys Gly Ile	
20 25 30	

Ile	Ala	Val 35	Ser	Arg	Gly	Gly	Leu 40	Val	Pro	Gly	Ala	Leu 45	Leu	Ala	Arg
Glu	Leu	Gly 50	Ile	Arg	His	Val 55	Asp	Thr	Val	Cys	Ile 60	Ser	Ser	Tyr	Asp
His 65	Asp	Asn	Gln	Arg	Glu 70	Leu	Lys	Val	Leu	Lys 75	Arg	Ala	Glu	Gly	Asp 80
Gly	Glu	Gly	Phe	Ile 85	Val	Ile	Asp	Asp	Leu 90	Val	Asp	Thr	Gly	Gly 95	Thr
Ala	Val	Ala 100	Ile	Arg	Glu	Met	Tyr	Pro 105	Lys	Ala	His	Phe	Val 110	Thr	Ile
Phe	Ala	Lys 115	Pro	Ala	Gly	Arg	Pro 120	Leu	Val	Asp	Asp	Tyr 125	Val	Val	Asp
Ile	Pro 130	Gln	Asp	Thr	Trp	Ile 135	Glu	Gln	Pro	Trp	Asp 140	Met	Gly	Val	Val
Phe 145	Val	Pro	Pro	Ile	Ser 150	Gly	Arg								

```
<210> 21
<211> 459
<212> DNA
<213> Escherichia coli
<220>
<221> CDS
<222> (1)..(456)
<223> coding for xanthine-guanine phosphoribosyl
transferase (gpt)
```

<400> 21																
atg	agc	gaa	aaa	tac	atc	gtc	acc	tgg	gac	atg	ttg	cag	atc	cat	gca	48
Met	Ser	Glu	Lys	Tyr	Ile	Val	Thr	Trp	Asp	Met	Leu	Gln	Ile	His	Ala	
1				5				10				15				
cgt	aaa	ctc	gca	agc	cga	ctg	atg	cct	tct	gaa	caa	tgg	aaa	ggc	att	96
Arg	Lys	Leu	Ala	Ser	Arg	Leu	Met	Pro	Ser	Glu	Gln	Trp	Lys	Gly	Ile	
				20				25				30				
att	gcc	gta	agc	cgt	ggc	ggg	ctg	gta	ccg	ggg	gcg	tta	ctg	gcg	cgt	144
Ile	Ala	Val	Ser	Arg	Gly	Gly	Leu	Val	Pro	Gly	Ala	Leu	Leu	Ala	Arg	
				35				40				45				
gaa	ctg	ggg	att	cgt	cat	gtc	gat	acc	gtt	tgt	att	tcc	agc	tac	gat	192
Glu	Leu	Gly	Ile	Arg	His	Val	Asp	Thr	Val	Cys	Ile	Ser	Ser	Tyr	Asp	
				50				55				60				
cac	gac	aac	cag	cgc	gag	ctt	aaa	gtg	ctg	aaa	cgc	gca	gaa	ggc	gat	240
His	Asp	Asn	Gln	Arg	Glu	Leu	Lys	Val	Leu	Lys	Arg	Ala	Glu	Gly	Asp	
65				70				75				80				
ggc	gaa	ggc	ttc	atc	gtt	att	gat	gac	ctg	gtg	gat	acc	ggg	ggg	act	288
Gly	Glu	Gly	Phe	Ile	Val	Ile	Asp	Asp	Leu	Val	Asp	Thr	Gly	Gly	Thr	
				85				90				95				
gcg	gtt	gcg	att	cgt	gaa	atg	tat	cca	aaa	gcg	cac	ttt	gtc	acc	atc	336
Ala	Val	Ala	Ile	Arg	Glu	Met	Tyr	Pro	Lys	Ala	His	Phe	Val	Thr	Ile	
				100				105				110				
ttc	gca	aaa	ccg	gct	ggg	cgt	ccg	ctg	gtt	gat	gac	tat	gtt	gtt	gat	384
Phe	Ala	Lys	Pro	Ala	Gly	Arg	Pro	Leu	Val	Asp	Asp	Tyr	Val	Val	Asp	
115				120				125								

atc ccg caa gat acc tgg att gaa cag ccg tgg gat atg ggc gtc gta 432  
 ile pro gln asp thr trp ile glu gln pro trp asp met gly val val  
 130 135 140

ttc gtc ccg cca atc tcc ggt cgc taa 459  
 phe val pro pro ile ser gly arg  
 145 150

<210> 22

<211> 152

<212> PRT

<213> Escherichia coli

<400> 22

Met Ser Glu Lys Tyr Ile Val Thr Trp Asp Met Leu Gln Ile His Ala  
 1 5 10 15

Arg Lys Leu Ala Ser Arg Leu Met Pro Ser Glu Gln Trp Lys Gly Ile  
 20 25 30

Ile Ala Val Ser Arg Gly Gly Leu Val Pro Gly Ala Leu Leu Ala Arg  
 35 40 45

Glu Leu Gly Ile Arg His Val Asp Thr Val Cys Ile Ser Ser Tyr Asp  
 50 55 60

His Asp Asn Gln Arg Glu Leu Lys Val Leu Lys Arg Ala Glu Gly Asp  
 65 70 75 80

Gly Glu Gly Phe Ile Val Ile Asp Asp Leu Val Asp Thr Gly Gly Thr  
 85 90 95

Ala Val Ala Ile Arg Glu Met Tyr Pro Lys Ala His Phe Val Thr Ile  
 100 105 110

Phe Ala Lys Pro Ala Gly Arg Pro Leu Val Asp Asp Tyr Val Val Asp  
 115 120 125

Ile Pro Gln Asp Thr Trp Ile Glu Gln Pro Trp Asp Met Gly Val Val  
 130 135 140

Phe Val Pro Pro Ile Ser Gly Arg  
 145 150

<210> 23

<211> 720

<212> DNA

<213> Escherichia coli

<220>

<221> CDS

<222> (1)..(717)

<223> coding for purine nucleoside phosphorylase (deoD)

<400> 23

atg gct acc cca cac att aat gca gaa atg ggc gat ttc gct gac gta 48  
 Met Ala Thr Pro His Ile Asn Ala Glu Met Gly Asp Phe Ala Asp Val  
 1 5 10 15

gtt ttg atg cca ggc gac ccg ctg cgt gcg aag tat att gct gaa act 96  
 Val Leu Met Pro Gly Asp Pro Leu Arg Ala Lys Tyr Ile Ala Glu Thr  
 20 25 30

ttc ctt gaa gat gcc cgt gaa gtg aac aac gtt cgc ggt atg ctg ggc 144  
 Phe Leu Glu Asp Ala Arg Glu Val Asn Asn Val Arg Gly Met Leu Gly  
 35 40 45

```

ttc acc ggt act tac aaa ggc cgc aaa att tcc gta atg ggt cac ggt 192
Phe Thr Gly Thr Tyr Lys Gly Arg Lys Ile Ser Val Met Gly His Gly
    50                55                60

atg ggt atc ccg tcc tgc tcc atc tac acc aaa gaa ctg atc acc gat 240
Met Gly Ile Pro Ser Cys Ser Ile Tyr Thr Lys Glu Leu Ile Thr Asp
    65                70                75                80

ttc ggc gtg aag aaa att atc cgc gtg ggt tcc tgt ggc gca gtt ctg 288
Phe Gly Val Lys Lys Ile Ile Arg Val Gly Ser Cys Gly Ala Val Leu
    85                90                95

ccg cac gta aaa ctg cgc gac gtc gtt atc ggt atg ggt gcc tgc acc 336
Pro His Val Lys Leu Arg Asp Val Val Ile Gly Met Gly Ala Cys Thr
    100               105               110

gat tcc aaa gtt aac cgc atc cgt ttt aaa gac cat gac ttt gcc gct 384
Asp Ser Lys Val Asn Arg Ile Arg Phe Lys Asp His Asp Phe Ala Ala
    115               120               125

atc gct gac ttc gac atg gtg cgt aac gca gta gat gca gct aaa gca 432
Ile Ala Asp Phe Asp Met Val Arg Asn Ala Val Asp Ala Ala Lys Ala
    130               135               140

ctg ggt att gat gct cgc gtg ggt aac ctg ttc tcc gct gac ctg ttc 480
Leu Gly Ile Asp Ala Arg Val Gly Asn Leu Phe Ser Ala Asp Leu Phe
    145               150               155               160

tac tct ccg gac ggc gaa atg ttc gac gtg atg gaa aaa tac ggc att 528
Tyr Ser Pro Asp Gly Glu Met Phe Asp Val Met Glu Lys Tyr Gly Ile
    165               170               175

ctc ggc gtg gaa atg gaa gcg gct ggt atc tac ggc gtc gct gca gaa 576
Leu Gly Val Glu Met Glu Ala Ala Gly Ile Tyr Gly Val Ala Ala Glu
    180               185               190

ttt ggc gcg aaa gcc ctg acc atc tgc acc gta tct gac cac atc cgc 624
Phe Gly Ala Lys Ala Leu Thr Ile Cys Thr Val Ser Asp His Ile Arg
    195               200               205

act cac gag cag acc act gcc gct gag cgt cag act acc ttc aac gac 672
Thr His Glu Gln Thr Thr Ala Ala Glu Arg Gln Thr Thr Phe Asn Asp
    210               215               220

atg atc aaa atc gca ctg gaa tcc gtt ctg ctg ggc gat aaa gag taa 720
Met Ile Lys Ile Ala Leu Glu Ser Val Leu Leu Gly Asp Lys Glu
    225               230               235

```

&lt;210&gt; 24

&lt;211&gt; 239

&lt;212&gt; PRT

&lt;213&gt; Escherichia coli

&lt;400&gt; 24

```

Met Ala Thr Pro His Ile Asn Ala Glu Met Gly Asp Phe Ala Asp Val
  1                5                10                15

Val Leu Met Pro Gly Asp Pro Leu Arg Ala Lys Tyr Ile Ala Glu Thr
    20                25                30

Phe Leu Glu Asp Ala Arg Glu Val Asn Asn Val Arg Gly Met Leu Gly
    35                40                45

Phe Thr Gly Thr Tyr Lys Gly Arg Lys Ile Ser Val Met Gly His Gly
    50                55                60

Met Gly Ile Pro Ser Cys Ser Ile Tyr Thr Lys Glu Leu Ile Thr Asp
    65                70                75                80

```



Phe Gly Val Lys Lys Ile Ile Arg Val Gly Ser Cys Gly Ala Val Leu  
                     85                    90                    95  
 Pro His Val Lys Leu Arg Asp Val Val Ile Gly Met Gly Ala Cys Thr  
                     100                    105                    110  
 Asp Ser Lys Val Asn Arg Ile Arg Phe Lys Asp His Asp Phe Ala Ala  
                     115                    120                    125  
 Ile Ala Asp Phe Asp Met Val Arg Asn Ala Val Asp Ala Ala Lys Ala  
                     130                    135                    140  
 Leu Gly Ile Asp Ala Arg Val Gly Asn Leu Phe Ser Ala Asp Leu Phe  
                     145                    150                    155                    160  
 Tyr Ser Pro Asp Gly Glu Met Phe Asp Val Met Glu Lys Tyr Gly Ile  
                     165                    170                    175  
 Leu Gly Val Glu Met Glu Ala Ala Gly Ile Tyr Gly Val Ala Ala Glu  
                     180                    185                    190  
 Phe Gly Ala Lys Ala Leu Thr Ile Cys Thr Val Ser Asp His Ile Arg  
                     195                    200                    205  
 Thr His Glu Gln Thr Thr Ala Ala Glu Arg Gln Thr Thr Phe Asn Asp  
                     210                    215                    220  
 Met Ile Lys Ile Ala Leu Glu Ser Val Leu Leu Gly Asp Lys Glu  
                     225                    230                    235

&lt;210&gt; 25

&lt;211&gt; 1545

&lt;212&gt; DNA

&lt;213&gt; Burkholderia caryophylli

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1542)

&lt;223&gt; coding for phosphonate monoester hydrolase (pehA)

&lt;400&gt; 25

atg acc aga aaa aat gtc ctg ctt atc gtc gtt gat caa tgg cga gca 48  
 Met Thr Arg Lys Asn Val Leu Leu Ile Val Val Asp Gln Trp Arg Ala  
   1                    5                    10                    15  
 gat ttt atc cct cac ctg atg cgg gcg gag ggg cgc gaa cct ttc ctt 96  
 Asp Phe Ile Pro His Leu Met Arg Ala Glu Gly Arg Glu Pro Phe Leu  
                     20                    25                    30  
 aaa act ccc aat ctt gat cgt ctt tgc cgg gaa ggc ttg acc ttc cgc 144  
 Lys Thr Pro Asn Leu Asp Arg Leu Cys Arg Glu Gly Leu Thr Phe Arg  
                     35                    40                    45  
 aat cat gtc acg acg tgc gtg ccg tgt ggt ccg gca agg gca agc ctg 192  
 Asn His Val Thr Thr Cys Val Pro Cys Gly Pro Ala Arg Ala Ser Leu  
                     50                    55                    60  
 ctg acg ggc ctc tac ctg atg aac cac cgg gcg gtg cag aac act gtt 240  
 Leu Thr Gly Leu Tyr Leu Met Asn His Arg Ala Val Gln Asn Thr Val  
                     65                    70                    75                    80  
 ccg ctt gac cag cgc cat cta aac ctt ggc aag gcc ctg cgc gcc att 288  
 Pro Leu Asp Gln Arg His Leu Asn Leu Gly Lys Ala Leu Arg Ala Ile  
                     85                    90                    95  
 ggc tac gat ccc gcg ctc att ggt tac acc acc acg aca cct gat ccg 336  
 Gly Tyr Asp Pro Ala Leu Ile Gly Tyr Thr Thr Thr Thr Pro Asp Pro  
                     100                    105                    110

cgc	aca	acc	tct	gca	agg	gat	ccg	cgt	ttc	acg	gtc	ctg	ggc	gac	atc	384
Arg	Thr	Thr	Ser	Ala	Arg	Asp	Pro	Arg	Phe	Thr	Val	Leu	Gly	Asp	Ile	
		115					120					125				
atg	gac	ggc	ttt	cgt	tcg	gtc	ggc	gca	ttc	gag	ccc	aat	atg	gag	ggg	432
Met	Asp	Gly	Phe	Arg	Ser	Val	Gly	Ala	Phe	Glu	Pro	Asn	Met	Glu	Gly	
	130					135					140					
tat	ttt	ggc	tgg	gtg	gcg	cag	aac	ggc	ttc	gaa	ctg	cca	gag	aac	cgc	480
Tyr	Phe	Gly	Trp	Val	Ala	Gln	Asn	Gly	Phe	Glu	Leu	Pro	Glu	Asn	Arg	
145					150					155					160	
gaa	gat	atc	tgg	ctg	ccg	gaa	ggg	gaa	cat	tcc	gtt	ccc	ggg	gct	acc	528
Glu	Asp	Ile	Trp	Leu	Pro	Glu	Gly	Glu	His	Ser	Val	Pro	Gly	Ala	Thr	
			165						170						175	
gac	aaa	ccg	tcg	cgc	att	ccg	aag	gaa	ttt	tcg	gat	tcg	aca	ttc	ttc	576
Asp	Lys	Pro	Ser	Arg	Ile	Pro	Lys	Glu	Phe	Ser	Asp	Ser	Thr	Phe	Phe	
		180						185						190		
acg	gag	cgc	gcc	ctg	aca	tat	ctg	aag	ggc	agg	gac	ggc	aag	cct	ttc	624
Thr	Glu	Arg	Ala	Leu	Thr	Tyr	Leu	Lys	Gly	Arg	Asp	Gly	Lys	Pro	Phe	
	195						200					205				
ttc	ctg	cat	ctt	ggc	tat	tat	cgc	ccg	cat	ccg	cct	ttc	gta	gcc	tcc	672
Phe	Leu	His	Leu	Gly	Tyr	Tyr	Arg	Pro	His	Pro	Pro	Phe	Val	Ala	Ser	
	210					215					220					
gcg	ccc	tac	cat	gcg	atg	tac	aaa	gcc	gaa	gat	atg	cct	gcg	cct	ata	720
Ala	Pro	Tyr	His	Ala	Met	Tyr	Lys	Ala	Glu	Asp	Met	Pro	Ala	Pro	Ile	
225					230					235					240	
cgt	gcg	gag	aat	ccg	gat	gcc	gaa	gcg	gca	cag	cat	ccg	ctc	atg	aag	768
Arg	Ala	Glu	Asn	Pro	Asp	Ala	Glu	Ala	Ala	Gln	His	Pro	Leu	Met	Lys	
			245					250						255		
cac	tat	atc	gac	cac	atc	aga	cgc	ggc	tcg	ttc	ttc	cat	ggc	gcg	gaa	816
His	Tyr	Ile	Asp	His	Ile	Arg	Arg	Gly	Ser	Phe	Phe	His	Gly	Ala	Glu	
		260						265					270			
ggc	tcg	gga	gca	acg	ctt	gat	gaa	ggc	gaa	att	cgc	cag	atg	cgc	gct	864
Gly	Ser	Gly	Ala	Thr	Leu	Asp	Glu	Gly	Glu	Ile	Arg	Gln	Met	Arg	Ala	
	275						280					285				
aca	tat	tgc	gga	ctg	atc	acc	gag	atc	gac	gat	tgt	ctg	ggg	agg	gtc	912
Thr	Tyr	Cys	Gly	Leu	Ile	Thr	Glu	Ile	Asp	Asp	Cys	Leu	Gly	Arg	Val	
	290					295					300					
ttt	gcc	tat	ctc	gat	gaa	acc	ggg	cag	tgg	gac	gac	acg	ctg	att	atc	960
Phe	Ala	Tyr	Leu	Asp	Glu	Thr	Gly	Gln	Trp	Asp	Asp	Thr	Leu	Ile	Ile	
305					310					315					320	
ttc	acg	agc	gat	cat	ggc	gaa	caa	ctg	ggc	gat	cat	cac	ctg	ctc	ggc	1008
Phe	Thr	Ser	Asp	His	Gly	Glu	Gln	Leu	Gly	Asp	His	His	Leu	Leu	Gly	
				325					330					335		
aag	atc	ggg	tac	aat	gcc	gaa	agc	ttc	cgt	att	ccc	ttg	gtc	ata	aag	1056
Lys	Ile	Gly	Tyr	Asn	Ala	Glu	Ser	Phe	Arg	Ile	Pro	Leu	Val	Ile	Lys	
		340						345					350			
gat	gcg	gga	cag	aac	ccg	cac	gcc	ggc	cag	atc	gaa	gaa	ggc	ttc	tcc	1104
Asp	Ala	Gly	Gln	Asn	Arg	His	Ala	Gly	Gln	Ile	Glu	Glu	Gly	Phe	Ser	
		355					360						365			
gaa	agc	atc	gac	gtc	atg	ccg	acc	atc	ctc	gaa	tgg	ctg	ggc	ggg	gaa	1152
Glu	Ser	Ile	Asp	Val	Met	Pro	Thr	Ile	Leu	Glu	Trp	Leu	Gly	Gly	Glu	
	370					375					380					

acg cct cgc gcc tgc gac ggc cgt tcg ctg ttg ccg ttt ctg gct gag 1200  
 Thr Pro Arg Ala Cys Asp Gly Arg Ser Leu Leu Pro Phe Leu Ala Glu  
 385 390 395 400  
 gga aag ccc tcc gac tgg cgc acg gaa cta cat tac gag ttc gat ttt 1248  
 Gly Lys Pro Ser Asp Trp Arg Thr Glu Leu His Tyr Glu Phe Asp Phe  
 405 410 415  
 cgc gat gtc ttc tac gat cag ccg cag aac tcg gtc cag ctt tcc cag 1296  
 Arg Asp Val Phe Tyr Asp Gln Pro Gln Asn Ser Val Gln Leu Ser Gln  
 420 425 430  
 gat gat tgc agc ctc tgt gtg atc gag gac gaa aac tac aag tac gtg 1344  
 Asp Asp Cys Ser Leu Cys Val Ile Glu Asp Glu Asn Tyr Lys Tyr Val  
 435 440 445  
 cat ttt gcc gcc ctg ccg ccg ctg ttc ttc gat ctg aag gca gac ccg 1392  
 His Phe Ala Ala Leu Pro Pro Leu Phe Phe Asp Leu Lys Ala Asp Pro  
 450 455 460  
 cat gaa ttc agc aat ctg gct ggc gat cct gct tat gcg gcc ctc gtt 1440  
 His Glu Phe Ser Asn Leu Ala Gly Asp Pro Ala Tyr Ala Ala Leu Val  
 465 470 475 480  
 cgt gac tat gcc cag aag gca ttg tcg tgg cga ctg tct cat gcc gac 1488  
 Arg Asp Tyr Ala Gln Lys Ala Leu Ser Trp Arg Leu Ser His Ala Asp  
 485 490 495  
 cgg aca ctc acc cat tac aga tcc agc ccg caa ggg ctg aca acg cgc 1536  
 Arg Thr Leu Thr His Tyr Arg Ser Ser Pro Gln Gly Leu Thr Thr Arg  
 500 505 510  
 aac cat tga 1545  
 Asn His  
 <210> 26  
 <211> 514  
 <212> PRT  
 <213> Burkholderia caryophylli  
 <400> 26  
 Met Thr Arg Lys Asn Val Leu Leu Ile Val Val Asp Gln Trp Arg Ala  
 1 5 10 15  
 Asp Phe Ile Pro His Leu Met Arg Ala Glu Gly Arg Glu Pro Phe Leu  
 20 25 30  
 Lys Thr Pro Asn Leu Asp Arg Leu Cys Arg Glu Gly Leu Thr Phe Arg  
 35 40 45  
 Asn His Val Thr Thr Cys Val Pro Cys Gly Pro Ala Arg Ala Ser Leu  
 50 55 60  
 Leu Thr Gly Leu Tyr Leu Met Asn His Arg Ala Val Gln Asn Thr Val  
 65 70 75 80  
 Pro Leu Asp Gln Arg His Leu Asn Leu Gly Lys Ala Leu Arg Ala Ile  
 85 90 95  
 Gly Tyr Asp Pro Ala Leu Ile Gly Tyr Thr Thr Thr Thr Pro Asp Pro  
 100 105 110  
 Arg Thr Thr Ser Ala Arg Asp Pro Arg Phe Thr Val Leu Gly Asp Ile  
 115 120 125  
 Met Asp Gly Phe Arg Ser Val Gly Ala Phe Glu Pro Asn Met Glu Gly  
 130 135 140

Tyr Phe Gly Trp Val Ala Gln Asn Gly Phe Glu Leu Pro Glu Asn Arg  
 145 150 155 160  
 Glu Asp Ile Trp Leu Pro Glu Gly Glu His Ser Val Pro Gly Ala Thr  
 165 170 175  
 Asp Lys Pro Ser Arg Ile Pro Lys Glu Phe Ser Asp Ser Thr Phe Phe  
 180 185 190  
 Thr Glu Arg Ala Leu Thr Tyr Leu Lys Gly Arg Asp Gly Lys Pro Phe  
 195 200 205  
 Phe Leu His Leu Gly Tyr Tyr Arg Pro His Pro Pro Phe Val Ala Ser  
 210 215 220  
 Ala Pro Tyr His Ala Met Tyr Lys Ala Glu Asp Met Pro Ala Pro Ile  
 225 230 235 240  
 Arg Ala Glu Asn Pro Asp Ala Glu Ala Ala Gln His Pro Leu Met Lys  
 245 250 255  
 His Tyr Ile Asp His Ile Arg Arg Gly Ser Phe Phe His Gly Ala Glu  
 260 265 270  
 Gly Ser Gly Ala Thr Leu Asp Glu Gly Glu Ile Arg Gln Met Arg Ala  
 275 280 285  
 Thr Tyr Cys Gly Leu Ile Thr Glu Ile Asp Asp Cys Leu Gly Arg Val  
 290 295 300  
 Phe Ala Tyr Leu Asp Glu Thr Gly Gln Trp Asp Asp Thr Leu Ile Ile  
 305 310 315 320  
 Phe Thr Ser Asp His Gly Glu Gln Leu Gly Asp His His Leu Leu Gly  
 325 330 335  
 Lys Ile Gly Tyr Asn Ala Glu Ser Phe Arg Ile Pro Leu Val Ile Lys  
 340 345 350  
 Asp Ala Gly Gln Asn Arg His Ala Gly Gln Ile Glu Glu Gly Phe Ser  
 355 360 365  
 Glu Ser Ile Asp Val Met Pro Thr Ile Leu Glu Trp Leu Gly Gly Glu  
 370 375 380  
 Thr Pro Arg Ala Cys Asp Gly Arg Ser Leu Leu Pro Phe Leu Ala Glu  
 385 390 395 400  
 Gly Lys Pro Ser Asp Trp Arg Thr Glu Leu His Tyr Glu Phe Asp Phe  
 405 410 415  
 Arg Asp Val Phe Tyr Asp Gln Pro Gln Asn Ser Val Gln Leu Ser Gln  
 420 425 430  
 Asp Asp Cys Ser Leu Cys Val Ile Glu Asp Glu Asn Tyr Lys Tyr Val  
 435 440 445  
 His Phe Ala Ala Leu Pro Pro Leu Phe Phe Asp Leu Lys Ala Asp Pro  
 450 455 460  
 His Glu Phe Ser Asn Leu Ala Gly Asp Pro Ala Tyr Ala Ala Leu Val  
 465 470 475 480  
 Arg Asp Tyr Ala Gln Lys Ala Leu Ser Trp Arg Leu Ser His Ala Asp  
 485 490 495  
 Arg Thr Leu Thr His Tyr Arg Ser Ser Pro Gln Gly Leu Thr Thr Arg  
 500 505 510  
 Asn His

```

<210> 27
<211> 2250
<212> DNA
<213> Agrobacterium rhizogenes
<220>
<221> CDS
<222> (1)..(2247)
<223> coding for tryptophane oxygenase (aux1)
<400> 27
atg gct gga tcc tcc ttc aca ttg cca tca act ggc tca gcg ccc ctt 48
Met Ala Gly Ser Ser Phe Thr Leu Pro Ser Thr Gly Ser Ala Pro Leu
1 5 10 15
gat atg atg ctt atc gat gat tca gat ctg ctg caa ttg ggt ctc cag 96
Asp Met Met Leu Ile Asp Asp Ser Asp Leu Leu Gln Leu Gly Leu Gln
20 25 30
cag gta ttc tcg aag cgg tac aca gag aca ccg cag tca cgc tac aaa 144
Gln Val Phe Ser Lys Arg Tyr Thr Glu Thr Pro Gln Ser Arg Tyr Lys
35 40 45
ctg acc agg agg gct tct cca gac gtc tca tct ggc gaa ggc aat gtg 192
Leu Thr Arg Arg Ala Ser Pro Asp Val Ser Ser Gly Glu Gly Asn Val
50 55 60
cat gcc ctt gcg ttc ata tat gtc aac gct gag acg ttg cag atg atc 240
His Ala Leu Ala Phe Ile Tyr Val Asn Ala Glu Thr Leu Gln Met Ile
65 70 75 80
aaa aac gct cga tcg cta acc gaa gcg aac ggc gtc aaa gat ctt gtc 288
Lys Asn Ala Arg Ser Leu Thr Glu Ala Asn Gly Val Lys Asp Leu Val
85 90 95
gcc atc gac gtt ccg cca ttt cga aac gac ttc tca aga gcg cta ctc 336
Ala Ile Asp Val Pro Pro Phe Arg Asn Asp Phe Ser Arg Ala Leu Leu
100 105 110
ctt caa gtg atc aac ttg ttg gga aac aac cga aat gcc gat gac gat 384
Leu Gln Val Ile Asn Leu Leu Gly Asn Asn Arg Asn Ala Asp Asp Asp
115 120 125
ctt agt cac ttc ata gca gtt gct ctc cca aac agc gcc cgc tct aag 432
Leu Ser His Phe Ile Ala Val Ala Leu Pro Asn Ser Ala Arg Ser Lys
130 135 140
atc cta acc acg gca ccg ttc gaa gga agc ttg tca gaa aac ttc agg 480
Ile Leu Thr Thr Ala Pro Phe Glu Gly Ser Leu Ser Glu Asn Phe Arg
145 150 155 160
ggg ttc ccg atc act cgt gaa gga aat gtg gca tgt gaa gtg cta gcc 528
Gly Phe Pro Ile Thr Arg Glu Gly Asn Val Ala Cys Glu Val Leu Ala
165 170 175
tat ggg aat aac ttg atg ccc aag gcc tgc tcc gat tcc ttt cca acc 576
Tyr Gly Asn Asn Leu Met Pro Lys Ala Cys Ser Asp Ser Phe Pro Thr
180 185 190
gtg gat ctt ctt tat gac tat ggc aag ttc ttc gag agt tgc gcg gcc 624
Val Asp Leu Leu Tyr Asp Tyr Gly Lys Phe Phe Glu Ser Cys Ala Ala
195 200 205
gat gga cgt atc ggt tat ttt cct gaa ggc gtt acg aaa cct aaa gtg 672
Asp Gly Arg Ile Gly Tyr Phe Pro Glu Gly Val Thr Lys Pro Lys Val
210 215 220

```

gct ata att ggc gca ggc ttt tcc ggg ctc gtt gca gcg agc gaa cta	720
Ala Ile Ile Gly Ala Gly Phe Ser Gly Leu Val Ala Ala Ser Glu Leu	
225 230 235 240	
ctt cat gca ggg gta gac gat gtt acg gtg tat gag gcg agt gat cgg	768
Leu His Ala Gly Val Asp Asp Val Thr Val Tyr Glu Ala Ser Asp Arg	
245 250 255	
ctt gga gga aag cta tgg tca cac gga ttt aag agt gct cca aat gtg	816
Leu Gly Gly Lys Leu Trp Ser His Gly Phe Lys Ser Ala Pro Asn Val	
260 265 270	
ata gcc gag atg ggg gcc atg cgt ttt ccg cga agt gaa tca tgc ttg	864
Ile Ala Glu Met Gly Ala Met Arg Phe Pro Arg Ser Glu Ser Cys Leu	
275 280 285	
ttc ttc tat ctc aaa aag cac gga ctg gac tcc gtt ggt ctg ttc ccg	912
Phe Phe Tyr Leu Lys Lys His Gly Leu Asp Ser Val Gly Leu Phe Pro	
290 295 300	
aat ccg gga agt gtc gat acc gca ttg ttc tac agg ggc cgt caa tat	960
Asn Pro Gly Ser Val Asp Thr Ala Leu Phe Tyr Arg Gly Arg Gln Tyr	
305 310 315 320	
atc tgg aaa gcg gga gag gag cca ccg gag ctg ttt cgt cgt gtg cac	1008
Ile Trp Lys Ala Gly Glu Glu Pro Pro Glu Leu Phe Arg Arg Val His	
325 330 335	
cat gga tgg cgc gca ttt ttg caa gat ggc tat ctc cat gat gga gtc	1056
His Gly Trp Arg Ala Phe Leu Gln Asp Gly Tyr Leu His Asp Gly Val	
340 345 350	
atg ttg gcg tca ccg tta gca att gtt gac gcc ttg aat tta ggg cat	1104
Met Leu Ala Ser Pro Leu Ala Ile Val Asp Ala Leu Asn Leu Gly His	
355 360 365	
cta cag cag gcg cat ggc ttc tgg caa tct tgg ctc aca tat ttt gag	1152
Leu Gln Gln Ala His Gly Phe Trp Gln Ser Trp Leu Thr Tyr Phe Glu	
370 375 380	
cga gag tct ttc tct tct ggc atc gaa aaa atg ttc ttg ggc aat cat	1200
Arg Glu Ser Phe Ser Ser Gly Ile Glu Lys Met Phe Leu Gly Asn His	
385 390 395 400	
cct ccg ggg ggt gaa caa tgg aat tcc cta gat gac ttg gat ctt ttc	1248
Pro Pro Gly Gly Glu Gln Trp Asn Ser Leu Asp Asp Leu Asp Leu Phe	
405 410 415	
aaa gcg ctg ggt att gga tcc ggc gga ttc ggc cct gta ttt gaa agt	1296
Lys Ala Leu Gly Ile Gly Ser Gly Gly Phe Gly Pro Val Phe Glu Ser	
420 425 430	
ggg ttt atc gag atc ctt cgc tta gtc gtc aac ggg tat gag gat aac	1344
Gly Phe Ile Glu Ile Leu Arg Leu Val Val Asn Gly Tyr Glu Asp Asn	
435 440 445	
gtg cgg ctg agt tac gaa gga att tct gag ctg cct cat agg atc gcc	1392
Val Arg Leu Ser Tyr Glu Gly Ile Ser Glu Leu Pro His Arg Ile Ala	
450 455 460	
tca cag gta att aac ggc aga tct att cgc gag cgt aca att cac gtt	1440
Ser Gln Val Ile Asn Gly Arg Ser Ile Arg Glu Arg Thr Ile His Val	
465 470 475 480	
caa gtc gag cag att gat aga gag gag gat aaa ata aat atc aag atc	1488
Gln Val Glu Gln Ile Asp Arg Glu Glu Asp Lys Ile Asn Ile Lys Ile	
485 490 495	

aaa gga gga aag gtt gag gtc tat gat cga gta ctg gtt aca tcc ggg	1536
Lys Gly Gly Lys Val Glu Val Tyr Asp Arg Val Leu Val Thr Ser Gly	
500 505 510	
ttt gcg aac atc gaa atg cgc cat ctc ctg aca tca agc aac gca ttc	1584
Phe Ala Asn Ile Glu Met Arg His Leu Leu Thr Ser Ser Asn Ala Phe	
515 520 525	
ttc cat gca gat gta agc cat gca ata ggg aac agt cat atg act ggt	1632
Phe His Ala Asp Val Ser His Ala Ile Gly Asn Ser His Met Thr Gly	
530 535 540	
gcg tca aaa ctg ttc ttg ctg act aac gaa aaa ttc tgg cta caa cat	1680
Ala Ser Lys Leu Phe Leu Leu Thr Asn Glu Lys Phe Trp Leu Gln His	
545 550 555 560	
cat ttg cca tcg tgc ata ctc acc acc ggc gtt gca aag gca gtt tat	1728
His Leu Pro Ser Cys Ile Leu Thr Thr Gly Val Ala Lys Ala Val Tyr	
565 570 575	
tgc tta gac tat gat ccg cga gat cca agc ggc aaa gga ctg gtg ttg	1776
Cys Leu Asp Tyr Asp Pro Arg Asp Pro Ser Gly Lys Gly Leu Val Leu	
580 585 590	
ata agc tat act tgg gag gat gac tca cat aag ctc cta gcc gtc ccc	1824
Ile Ser Tyr Thr Trp Glu Asp Asp Ser His Lys Leu Leu Ala Val Pro	
595 600 605	
gac aaa aga gaa agg ttc gca tcg ctg cag cgc gat att ggg agg gca	1872
Asp Lys Arg Glu Arg Phe Ala Ser Leu Gln Arg Asp Ile Gly Arg Ala	
610 615 620	
ttc cca gat ttt gcc aag cac cta act cct gca gac ggg aac tat gat	1920
Phe Pro Asp Phe Ala Lys His Leu Thr Pro Ala Asp Gly Asn Tyr Asp	
625 630 635 640	
gat aat atc gtt caa cat gat tgg ctg act gat ccc cac gct ggc gga	1968
Asp Asn Ile Val Gln His Asp Trp Leu Thr Asp Pro His Ala Gly Gly	
645 650 655	
gcg ttt aaa ctg aac cgc aga ggc aac gac gta tat tca gaa agg ctt	2016
Ala Phe Lys Leu Asn Arg Arg Gly Asn Asp Val Tyr Ser Glu Arg Leu	
660 665 670	
ttc ttt cag ccc ttt gac gta atg cat ccc gcg gac gat aag gga ctt	2064
Phe Phe Gln Pro Phe Asp Val Met His Pro Ala Asp Asp Lys Gly Leu	
675 680 685	
tac ttg gcc ggt tgt agc tgt tcc ttc acc gga ggg tgg gtt cat ggt	2112
Tyr Leu Ala Gly Cys Ser Cys Ser Phe Thr Gly Gly Trp Val His Gly	
690 695 700	
gcc att cag acc gca tgc aac gct acg tgt gcg atc att tat ggt tcc	2160
Ala Ile Gln Thr Ala Cys Asn Ala Thr Cys Ala Ile Ile Tyr Gly Ser	
705 710 715 720	
gga cac ctg caa gag cta atc cac tgg cga cac ctc aaa gaa ggt aat	2208
Gly His Leu Gln Glu Leu Ile His Trp Arg His Leu Lys Glu Gly Asn	
725 730 735	
cca ctg gcg cac gct tgg aag cgg tat agg tat caa gcg tga	2250
Pro Leu Ala His Ala Trp Lys Arg Tyr Arg Tyr Gln Ala	
740 745	

&lt;210&gt; 28

&lt;211&gt; 749

&lt;212&gt; PRT

&lt;213&gt; Agrobacterium rhizogenes

&lt;400&gt; 28

```

Met Ala Gly Ser Ser Phe Thr Leu Pro Ser Thr Gly Ser Ala Pro Leu
 1              5              10              15
Asp Met Met Leu Ile Asp Asp Ser Asp Leu Leu Gln Leu Gly Leu Gln
      20              25              30
Gln Val Phe Ser Lys Arg Tyr Thr Glu Thr Pro Gln Ser Arg Tyr Lys
      35              40              45
Leu Thr Arg Arg Ala Ser Pro Asp Val Ser Ser Gly Glu Gly Asn Val
 50              55              60
His Ala Leu Ala Phe Ile Tyr Val Asn Ala Glu Thr Leu Gln Met Ile
 65              70              75              80
Lys Asn Ala Arg Ser Leu Thr Glu Ala Asn Gly Val Lys Asp Leu Val
      85              90              95
Ala Ile Asp Val Pro Pro Phe Arg Asn Asp Phe Ser Arg Ala Leu Leu
      100              105              110
Leu Gln Val Ile Asn Leu Leu Gly Asn Asn Arg Asn Ala Asp Asp Asp
      115              120              125
Leu Ser His Phe Ile Ala Val Ala Leu Pro Asn Ser Ala Arg Ser Lys
      130              135              140
Ile Leu Thr Thr Ala Pro Phe Glu Gly Ser Leu Ser Glu Asn Phe Arg
      145              150              155              160
Gly Phe Pro Ile Thr Arg Glu Gly Asn Val Ala Cys Glu Val Leu Ala
      165              170              175
Tyr Gly Asn Asn Leu Met Pro Lys Ala Cys Ser Asp Ser Phe Pro Thr
      180              185              190
Val Asp Leu Leu Tyr Asp Tyr Gly Lys Phe Phe Glu Ser Cys Ala Ala
      195              200              205
Asp Gly Arg Ile Gly Tyr Phe Pro Glu Gly Val Thr Lys Pro Lys Val
      210              215              220
Ala Ile Ile Gly Ala Gly Phe Ser Gly Leu Val Ala Ala Ser Glu Leu
      225              230              235              240
Leu His Ala Gly Val Asp Asp Val Thr Val Tyr Glu Ala Ser Asp Arg
      245              250              255
Leu Gly Gly Lys Leu Trp Ser His Gly Phe Lys Ser Ala Pro Asn Val
      260              265              270
Ile Ala Glu Met Gly Ala Met Arg Phe Pro Arg Ser Glu Ser Cys Leu
      275              280              285
Phe Phe Tyr Leu Lys Lys His Gly Leu Asp Ser Val Gly Leu Phe Pro
      290              295              300
Asn Pro Gly Ser Val Asp Thr Ala Leu Phe Tyr Arg Gly Arg Gln Tyr
      305              310              315              320
Ile Trp Lys Ala Gly Glu Glu Pro Pro Glu Leu Phe Arg Arg Val His
      325              330              335
His Gly Trp Arg Ala Phe Leu Gln Asp Gly Tyr Leu His Asp Gly Val
      340              345              350

```



Met Leu Ala Ser Pro Leu Ala Ile Val Asp Ala Leu Asn Leu Gly His  
 355 360 365  
 Leu Gln Gln Ala His Gly Phe Trp Gln Ser Trp Leu Thr Tyr Phe Glu  
 370 375 380  
 Arg Glu Ser Phe Ser Ser Gly Ile Glu Lys Met Phe Leu Gly Asn His  
 385 390 395 400  
 Pro Pro Gly Gly Glu Gln Trp Asn Ser Leu Asp Asp Leu Asp Leu Phe  
 405 410 415  
 Lys Ala Leu Gly Ile Gly Ser Gly Gly Phe Gly Pro Val Phe Glu Ser  
 420 425 430  
 Gly Phe Ile Glu Ile Leu Arg Leu Val Val Asn Gly Tyr Glu Asp Asn  
 435 440 445  
 Val Arg Leu Ser Tyr Glu Gly Ile Ser Glu Leu Pro His Arg Ile Ala  
 450 455 460  
 Ser Gln Val Ile Asn Gly Arg Ser Ile Arg Glu Arg Thr Ile His Val  
 465 470 475 480  
 Gln Val Glu Gln Ile Asp Arg Glu Glu Asp Lys Ile Asn Ile Lys Ile  
 485 490 495  
 Lys Gly Gly Lys Val Glu Val Tyr Asp Arg Val Leu Val Thr Ser Gly  
 500 505 510  
 Phe Ala Asn Ile Glu Met Arg His Leu Leu Thr Ser Ser Asn Ala Phe  
 515 520 525  
 Phe His Ala Asp Val Ser His Ala Ile Gly Asn Ser His Met Thr Gly  
 530 535 540  
 Ala Ser Lys Leu Phe Leu Leu Thr Asn Glu Lys Phe Trp Leu Gln His  
 545 550 555 560  
 His Leu Pro Ser Cys Ile Leu Thr Thr Gly Val Ala Lys Ala Val Tyr  
 565 570 575  
 Cys Leu Asp Tyr Asp Pro Arg Asp Pro Ser Gly Lys Gly Leu Val Leu  
 580 585 590  
 Ile Ser Tyr Thr Trp Glu Asp Asp Ser His Lys Leu Leu Ala Val Pro  
 595 600 605  
 Asp Lys Arg Glu Arg Phe Ala Ser Leu Gln Arg Asp Ile Gly Arg Ala  
 610 615 620  
 Phe Pro Asp Phe Ala Lys His Leu Thr Pro Ala Asp Gly Asn Tyr Asp  
 625 630 635 640  
 Asp Asn Ile Val Gln His Asp Trp Leu Thr Asp Pro His Ala Gly Gly  
 645 650 655  
 Ala Phe Lys Leu Asn Arg Arg Gly Asn Asp Val Tyr Ser Glu Arg Leu  
 660 665 670  
 Phe Phe Gln Pro Phe Asp Val Met His Pro Ala Asp Asp Lys Gly Leu  
 675 680 685  
 Tyr Leu Ala Gly Cys Ser Cys Ser Phe Thr Gly Gly Trp Val His Gly  
 690 695 700  
 Ala Ile Gln Thr Ala Cys Asn Ala Thr Cys Ala Ile Ile Tyr Gly Ser  
 705 710 715 720  
 Gly His Leu Gln Glu Leu Ile His Trp Arg His Leu Lys Glu Gly Asn  
 725 730 735

Pro Leu Ala His Ala Trp Lys Arg Tyr Arg Tyr Gln Ala  
 740 745

<210> 29

<211> 1401

<212> DNA

<213> Agrobacterium rhizogenes

<220>

<221> CDS

<222> (1)..(1398)

<223> coding for indole acetamide hydrolase

<400> 29

atg gtg acc ctc tcc tcg atc acc gag acg ctt aaa tgt ctc agg gaa	48
Met Val Thr Leu Ser Ser Ile Thr Glu Thr Leu Lys Cys Leu Arg Glu	
1 5 10 15	
aga aaa tac tcg tgc ttt gag tta atc gaa acg ata ata gcc cgc tgt	96
Arg Lys Tyr Ser Cys Phe Glu Leu Ile Glu Thr Ile Ile Ala Arg Cys	
20 25 30	
gaa gca gca aga tcc tta aac gcc ttt ctg gaa acc gac tgg gcg cac	144
Glu Ala Ala Arg Ser Leu Asn Ala Phe Leu Glu Thr Asp Trp Ala His	
35 40 45	
cta cgg tgg act gcc agc aaa atc gat caa cac gga ggt gcc ggt gtt	192
Leu Arg Trp Thr Ala Ser Lys Ile Asp Gln His Gly Gly Ala Gly Val	
50 55 60	
ggc cta gct ggc gtt ccc cta tgc ttt aaa gcg aat att gcg aca ggc	240
Gly Leu Ala Gly Val Pro Leu Cys Phe Lys Ala Asn Ile Ala Thr Gly	
65 70 75 80	
agg ttc gcc gcg acc gct ggt acg cca ggc tta cag aac cac aaa ccc	288
Arg Phe Ala Ala Thr Ala Gly Thr Pro Gly Leu Gln Asn His Lys Pro	
85 90 95	
aag acg cct gcc gga gtt gca cga caa ctt ctc gcg gct ggg gca ctg	336
Lys Thr Pro Ala Gly Val Ala Arg Gln Leu Leu Ala Ala Gly Ala Leu	
100 105 110	
cct ggc gct tcg gga aac atg cac gaa ttg tct ttt ggg atc acg agc	384
Pro Gly Ala Ser Gly Asn Met His Glu Leu Ser Phe Gly Ile Thr Ser	
115 120 125	
aac aac ttc gcc aca ggc gcc gta cga aac ccg tgg aac cct agt ctc	432
Asn Asn Phe Ala Thr Gly Ala Val Arg Asn Pro Trp Asn Pro Ser Leu	
130 135 140	
atc cca ggg gga tca agt ggg ggt gtg gcc gcc gcg gtg gcc ggc cga	480
Ile Pro Gly Gly Ser Ser Gly Gly Val Ala Ala Ala Val Ala Gly Arg	
145 150 155 160	
ttg atg ctg ggc ggc gtc gga act gac acg gga gcg tcg gtc cgt tta	528
Leu Met Leu Gly Gly Val Gly Thr Asp Thr Gly Ala Ser Val Arg Leu	
165 170 175	
ccg gcc gcc ttg tgc ggc gtg gtg ggg ttt cgt cct acc gtg ggg cga	576
Pro Ala Ala Leu Cys Gly Val Val Gly Phe Arg Pro Thr Val Gly Arg	
180 185 190	
tat cca acg gac gga ata gtt ccg gta agc ccc acc cgg gac acc cct	624
Tyr Pro Thr Asp Gly Ile Val Pro Val Ser Pro Thr Arg Asp Thr Pro	
195 200 205	

ggc gtt atc gca cag aat gtt ccg gac gtg att ctt ctt gac ggt atc	672
Gly Val Ile Ala Gln Asn Val Pro Asp Val Ile Leu Leu Asp Gly Ile	
210 215 220	
att tgc ggg aga ccg ccg gtt aat caa acg gtc cgc ctg aag ggg ctg	720
Ile Cys Gly Arg Pro Pro Val Asn Gln Thr Val Arg Leu Lys Gly Leu	
225 230 235 240	
cgt ata ggc ttg cca acc gct tac ttt tac aac gac ctg gag ccc gat	768
Arg Ile Gly Leu Pro Thr Ala Tyr Phe Tyr Asn Asp Leu Glu Pro Asp	
245 250 255	
gtc gcc tta gca gcc gag acg att atc aga gtt ctg gca cgc aaa gat	816
Val Ala Leu Ala Ala Glu Thr Ile Ile Arg Val Leu Ala Arg Lys Asp	
260 265 270	
gtt act ttt gtt gaa gca gat att cct gat tta gcg cat cac aat gaa	864
Val Thr Phe Val Glu Ala Asp Ile Pro Asp Leu Ala His His Asn Glu	
275 280 285	
ggg gtc agc ttt ccg act gcc atc tac gaa ttt ccg ttg tcc ctt gaa	912
Gly Val Ser Phe Pro Thr Ala Ile Tyr Glu Phe Pro Leu Ser Leu Glu	
290 295 300	
cat tat att cag aac ttc gta gag ggt gtt tcc ttt tct gag gtt gtc	960
His Tyr Ile Gln Asn Phe Val Glu Gly Val Ser Phe Ser Glu Val Val	
305 310 315 320	
aga gcg att cgc agt ccg gat gtt gca agt att ctc aat gca caa ctc	1008
Arg Ala Ile Arg Ser Pro Asp Val Ala Ser Ile Leu Asn Ala Gln Leu	
325 330 335	
tcg gat aat ctt att tcc aaa agc gag tat tgt ctg gcg cga cgt ttt	1056
Ser Asp Asn Leu Ile Ser Lys Ser Glu Tyr Cys Leu Ala Arg Arg Phe	
340 345 350	
ttc aga ccg aga ctc caa gcg gcc tac cac agt tac ttc aag gcg cat	1104
Phe Arg Pro Arg Leu Gln Ala Ala Tyr His Ser Tyr Phe Lys Ala His	
355 360 365	
cag cta gat gca att ctt ttc cca aca gct ccg ttg aca gcc aag cca	1152
Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Thr Ala Lys Pro	
370 375 380	
att ggc cat gat cta tcg gtg att cac aat ggc tca atg acc gat acc	1200
Ile Gly His Asp Leu Ser Val Ile His Asn Gly Ser Met Thr Asp Thr	
385 390 395 400	
ttt aaa atc ttc gtg cgg aat gta gat ccc agc agt aat gcg ggc ctg	1248
Phe Lys Ile Phe Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu	
405 410 415	
ccg ggc cta agt ctt ccc gtt tct ctt agt tcc aac ggt ctg cct att	1296
Pro Gly Leu Ser Leu Pro Val Ser Leu Ser Ser Asn Gly Leu Pro Ile	
420 425 430	
ggc atg gaa atc gat ggc tct gca agc tcg gat gaa cgt ctg tta gca	1344
Gly Met Glu Ile Asp Gly Ser Ala Ser Ser Asp Glu Arg Leu Leu Ala	
435 440 445	
att gga cta gcg ata gaa gaa gca ata gac ttt agg cat cgt ccg act	1392
Ile Gly Leu Ala Ile Glu Glu Ala Ile Asp Phe Arg His Arg Pro Thr	
450 455 460	
ctg tcg taa	1401
Leu Ser	
465	

&lt;210&gt; 30

&lt;211&gt; 466

&lt;212&gt; PRT

&lt;213&gt; Agrobacterium rhizogenes

&lt;400&gt; 30

```

Met Val Thr Leu Ser Ser Ile Thr Glu Thr Leu Lys Cys Leu Arg Glu
  1           5           10           15
Arg Lys Tyr Ser Cys Phe Glu Leu Ile Glu Thr Ile Ile Ala Arg Cys
          20           25           30
Glu Ala Ala Arg Ser Leu Asn Ala Phe Leu Glu Thr Asp Trp Ala His
          35           40           45
Leu Arg Trp Thr Ala Ser Lys Ile Asp Gln His Gly Gly Ala Gly Val
          50           55           60
Gly Leu Ala Gly Val Pro Leu Cys Phe Lys Ala Asn Ile Ala Thr Gly
          65           70           75           80
Arg Phe Ala Ala Thr Ala Gly Thr Pro Gly Leu Gln Asn His Lys Pro
          85           90           95
Lys Thr Pro Ala Gly Val Ala Arg Gln Leu Leu Ala Ala Gly Ala Leu
          100          105          110
Pro Gly Ala Ser Gly Asn Met His Glu Leu Ser Phe Gly Ile Thr Ser
          115          120          125
Asn Asn Phe Ala Thr Gly Ala Val Arg Asn Pro Trp Asn Pro Ser Leu
          130          135          140
Ile Pro Gly Gly Ser Ser Gly Gly Val Ala Ala Ala Val Ala Gly Arg
          145          150          155          160
Leu Met Leu Gly Gly Val Gly Thr Asp Thr Gly Ala Ser Val Arg Leu
          165          170          175
Pro Ala Ala Leu Cys Gly Val Val Gly Phe Arg Pro Thr Val Gly Arg
          180          185          190
Tyr Pro Thr Asp Gly Ile Val Pro Val Ser Pro Thr Arg Asp Thr Pro
          195          200          205
Gly Val Ile Ala Gln Asn Val Pro Asp Val Ile Leu Leu Asp Gly Ile
          210          215          220
Ile Cys Gly Arg Pro Pro Val Asn Gln Thr Val Arg Leu Lys Gly Leu
          225          230          235          240
Arg Ile Gly Leu Pro Thr Ala Tyr Phe Tyr Asn Asp Leu Glu Pro Asp
          245          250          255
Val Ala Leu Ala Ala Glu Thr Ile Ile Arg Val Leu Ala Arg Lys Asp
          260          265          270
Val Thr Phe Val Glu Ala Asp Ile Pro Asp Leu Ala His His Asn Glu
          275          280          285
Gly Val Ser Phe Pro Thr Ala Ile Tyr Glu Phe Pro Leu Ser Leu Glu
          290          295          300
His Tyr Ile Gln Asn Phe Val Glu Gly Val Ser Phe Ser Glu Val Val
          305          310          315          320
Arg Ala Ile Arg Ser Pro Asp Val Ala Ser Ile Leu Asn Ala Gln Leu
          325          330          335
Ser Asp Asn Leu Ile Ser Lys Ser Glu Tyr Cys Leu Ala Arg Arg Phe
          340          345          350

```

Phe Arg Pro Arg Leu Gln Ala Ala Tyr His Ser Tyr Phe Lys Ala His  
                   355                                  360                                  365  
 Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Thr Ala Lys Pro  
                   370                                  375                                  380  
 Ile Gly His Asp Leu Ser Val Ile His Asn Gly Ser Met Thr Asp Thr  
 385                                  390                                  395                                  400  
 Phe Lys Ile Phe Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu  
                                   405                                  410                                  415  
 Pro Gly Leu Ser Leu Pro Val Ser Leu Ser Ser Asn Gly Leu Pro Ile  
                                   420                                  425                                  430  
 Gly Met Glu Ile Asp Gly Ser Ala Ser Ser Asp Glu Arg Leu Leu Ala  
                   435                                  440                                  445  
 Ile Gly Leu Ala Ile Glu Glu Ala Ile Asp Phe Arg His Arg Pro Thr  
                   450                                  455                                  460  
 Leu Ser  
 465

<210> 31  
 <211> 2268  
 <212> DNA  
 <213> Agrobacterium tumefaciens  
 <220>  
 <221> CDS  
 <222> (1)..(2265)  
 <223> coding for tryptophan monooxygenase

<400> 31  
 atg tca gct tca cct ctc ctt gat aac cag tgc gat cat ttc tct acc 48  
 Met Ser Ala Ser Pro Leu Leu Asp Asn Gln Cys Asp His Phe Ser Thr  
   1                                  5                                  10                                  15  
 aaa atg gtg gat ctg ata atg gtc gat aag gct gat gaa ttg gac cgc 96  
 Lys Met Val Asp Leu Ile Met Val Asp Lys Ala Asp Glu Leu Asp Arg  
                   20                                  25                                  30  
 agg gtt tcc gat gcc ttc tca gaa cgt gaa gct tct agg gga agg agg 144  
 Arg Val Ser Asp Ala Phe Ser Glu Arg Glu Ala Ser Arg Gly Arg Arg  
                   35                                  40                                  45  
 att act caa atc tcc ggc gag tgc agc gct ggg tta gct tgc aaa agg 192  
 Ile Thr Gln Ile Ser Gly Glu Cys Ser Ala Gly Leu Ala Cys Lys Arg  
                   50                                  55                                  60  
 ctg gcc gac ggt cgc ttt ccc gag atc tca act ggt gag aag gta gca 240  
 Leu Ala Asp Gly Arg Phe Pro Glu Ile Ser Thr Gly Glu Lys Val Ala  
                   65                                  70                                  75                                  80  
 gcc ctc tcc gct tac atc tat gtt ggc aag gaa att ctg ggg cgg ata 288  
 Ala Leu Ser Ala Tyr Ile Tyr Val Gly Lys Glu Ile Leu Gly Arg Ile  
                                   85                                  90                                  95  
 ctt gaa tcg gaa cct tgg gcg cga gca aga gtg agt ggt ctc gtt gcc 336  
 Leu Glu Ser Glu Pro Trp Ala Arg Ala Arg Val Ser Gly Leu Val Ala  
                   100                                  105                                  110  
 atc gac ctt gca cca ttt tgt atg gat ttc tcc gaa gca caa ctt ctc 384  
 Ile Asp Leu Ala Pro Phe Cys Met Asp Phe Ser Glu Ala Gln Leu Leu  
                   115                                  120                                  125

caa acc ctg ttt ttg ctg agc ggt aaa aga tgt gca tcc agc gat ctt	432
Gln Thr Leu Phe Leu Leu Ser Gly Lys Arg Cys Ala Ser Ser Asp Leu	
130 135 140	
agt cat ttc gtg gcc att tca atc tct aag act gcc cgc tcc cga acc	480
Ser His Phe Val Ala Ile Ser Ile Ser Lys Thr Ala Arg Ser Arg Thr	
145 150 155 160	
ctg caa atg ccg ccg tac gag aaa ggc acg acg aaa cgc gtt acc ggg	528
Leu Gln Met Pro Tyr Glu Lys Gly Thr Thr Lys Arg Val Thr Gly	
165 170 175	
ttt acc ctg acc ctt gaa gag gcc gta cca ttt gac atg gta gct tat	576
Phe Thr Leu Thr Leu Glu Glu Ala Val Pro Phe Asp Met Val Ala Tyr	
180 185 190	
ggt cga aac ctg atg ctg aag gct tcg gca ggt tcc ttt cca aca att	624
Gly Arg Asn Leu Met Leu Lys Ala Ser Ala Gly Ser Phe Pro Thr Ile	
195 200 205	
gac ttg ctc tat gac tac aga tcg ttt ttt gac caa tgt tcc gat att	672
Asp Leu Leu Tyr Asp Tyr Arg Ser Phe Phe Asp Gln Cys Ser Asp Ile	
210 215 220	
gga cgg atc ggc ttc ttt ccg gaa gat gtt cct aag ccg aaa gtg gcg	720
Gly Arg Ile Gly Phe Phe Pro Glu Asp Val Pro Lys Pro Lys Val Ala	
225 230 235 240	
atc att ggc gct ggc att tcc gga ctc gtg gta gca agc gaa ctg ctt	768
Ile Ile Gly Ala Gly Ile Ser Gly Leu Val Val Ala Ser Glu Leu Leu	
245 250 255	
cat gct ggt gta gac gat gtt aca ata tat gaa gca agt gat cgg gtt	816
His Ala Gly Val Asp Asp Val Thr Ile Tyr Glu Ala Ser Asp Arg Val	
260 265 270	
gga ggc aag ctt tgg tca cat gct ttc aag gat gct ccc agc gtg gtg	864
Gly Gly Lys Leu Trp Ser His Ala Phe Lys Asp Ala Pro Ser Val Val	
275 280 285	
gcc gaa atg ggg gcg atg cga ttt cct cct gct gca tcg tgc ttg ttt	912
Ala Glu Met Gly Ala Met Arg Phe Pro Pro Ala Ala Ser Cys Leu Phe	
290 295 300	
ttc ttc ctc gag cgg tac ggc ctg tct tcg atg agg ccg ttc cca aat	960
Phe Phe Leu Glu Arg Tyr Gly Leu Ser Ser Met Arg Pro Phe Pro Asn	
305 310 315 320	
ccc ggc aca gtc gac act aac ttg gtc tac caa ggc ctc cga tac gtg	1008
Pro Gly Thr Val Asp Thr Asn Leu Val Tyr Gln Gly Leu Arg Tyr Val	
325 330 335	
tgg aaa gcc ggg cag cag cca ccg aag ctg ttc cat cgc gtt tac agc	1056
Trp Lys Ala Gly Gln Gln Pro Pro Lys Leu Phe His Arg Val Tyr Ser	
340 345 350	
ggt tgg cgt gcg ttc ttg agg gac ggt ttc cat gag gga gat att gtg	1104
Gly Trp Arg Ala Phe Leu Arg Asp Gly Phe His Glu Gly Asp Ile Val	
355 360 365	
ttg gct tcg cct gtt gtt att act caa gcc ttg aaa tca gga gac att	1152
Leu Ala Ser Pro Val Val Ile Thr Gln Ala Leu Lys Ser Gly Asp Ile	
370 375 380	
agg cgg gct cat gac tcc tgg caa act tgg ctg aac cgt ttc ggg agg	1200
Arg Arg Ala His Asp Ser Trp Gln Thr Trp Leu Asn Arg Phe Gly Arg	
385 390 395 400	

gag tcc ttc tct tca gcg ata gag agg atc ttt ctg ggc acg cat cct	1248
Glu Ser Phe Ser Ser Ala Ile Glu Arg Ile Phe Leu Gly Thr His Pro	
405 410 415	
cct ggt ggt gaa aca tgg agt ttc cct cat gat tgg gac cta ttc aag	1296
Pro Gly Gly Glu Thr Trp Ser Phe Pro His Asp Trp Asp Leu Phe Lys	
420 425 430	
cta atg gga ata gga tct ggc ggg ttt ggt cca gtt ttt gaa agc ggg	1344
Leu Met Gly Ile Gly Ser Gly Gly Phe Gly Pro Val Phe Glu Ser Gly	
435 440 445	
ttt att gag atc ctt cgc ttg gtc ata aac gga tat gaa gaa aat cag	1392
Phe Ile Glu Ile Leu Arg Leu Val Ile Asn Gly Tyr Glu Glu Asn Gln	
450 455 460	
cgg atg tgc tct gaa gga atc tca gaa ctt cca cgt cga ata gcc tct	1440
Arg Met Cys Ser Glu Gly Ile Ser Glu Leu Pro Arg Arg Ile Ala Ser	
465 470 475 480	
caa gtg gtt aac ggt gtg tct gta agc cag cgt ata cgc cat gtt caa	1488
Gln Val Val Asn Gly Val Ser Val Ser Gln Arg Ile Arg His Val Gln	
485 490 495	
gtc agg gcg att gag aag gaa aag aca aaa ata aag ata agg ctt aag	1536
Val Arg Ala Ile Glu Lys Glu Lys Thr Lys Ile Lys Ile Arg Leu Lys	
500 505 510	
agc ggg ata tct gaa ctt tat gat aag gtg gtg gtt aca tct gga ctc	1584
Ser Gly Ile Ser Glu Leu Tyr Asp Lys Val Val Val Thr Ser Gly Leu	
515 520 525	
gca aat atc caa ctc agg cat tgt ctg aca tgc gat acc acc att ttt	1632
Ala Asn Ile Gln Leu Arg His Cys Leu Thr Cys Asp Thr Thr Ile Phe	
530 535 540	
cgt gca cca gtg aac caa gcg gtt gat aac agc cat atg aca ggc tcg	1680
Arg Ala Pro Val Asn Gln Ala Val Asp Asn Ser His Met Thr Gly Ser	
545 550 555 560	
tca aaa ctc ttt ctg ctg act gaa cga aaa ttt tgg tta gac cat atc	1728
Ser Lys Leu Phe Leu Leu Thr Glu Arg Lys Phe Trp Leu Asp His Ile	
565 570 575	
ctc ccg tcc tgt gtc ctc atg gac ggg atc gca aaa gca gtg tac tgc	1776
Leu Pro Ser Cys Val Leu Met Asp Gly Ile Ala Lys Ala Val Tyr Cys	
580 585 590	
ttg gac tat gag ccg cag gat ccg aat ggt aaa ggt ctg gtg ccc ccc	1824
Leu Asp Tyr Glu Pro Gln Asp Pro Asn Gly Lys Gly Leu Val Pro Pro	
595 600 605	
act tat aca tgg gag gac gac tcc cac aag ctg ttg gcg gtt ccc gac	1872
Thr Tyr Thr Trp Glu Asp Asp Ser His Lys Leu Leu Ala Val Pro Asp	
610 615 620	
aaa aaa gag cga ttc tgt ctg ctg ccg gac gca att tcg aga tct ttc	1920
Lys Lys Glu Arg Phe Cys Leu Leu Arg Asp Ala Ile Ser Arg Ser Phe	
625 630 635 640	
ccg gcg ttt gcc cag cat cta gtt cct gcc tgc gct gat tac gac caa	1968
Pro Ala Phe Ala Gln His Leu Val Pro Ala Cys Ala Asp Tyr Asp Gln	
645 650 655	
aat gtt gtt caa cat gat tgg ctt aca gac gag aat gcc ggg gga gct	2016
Asn Val Val Gln His Asp Trp Leu Thr Asp Glu Asn Ala Gly Gly Ala	
660 665 670	

```

ttc aaa ctc aac cgg cgt ggc gag gat ttt tat tct gaa gaa ctt ttc 2064
Phe Lys Leu Asn Arg Arg Gly Glu Asp Phe Tyr Ser Glu Glu Leu Phe
      675                      680                      685

ttt caa gcg ctg gac atg cct aat gat acc gga gtt tac ttg gcg ggt 2112
Phe Gln Ala Leu Asp Met Pro Asn Asp Thr Gly Val Tyr Leu Ala Gly
      690                      695                      700

tgc agt tgt tcc ttc acc ggt gga tgg gtg gag ggc gct att cag acc 2160
Cys Ser Cys Ser Phe Thr Gly Gly Trp Val Glu Gly Ala Ile Gln Thr
      705                      710                      715                      720

gcg tgt aac gcc gtc tgt gca att atc cac aat tgt gga ggt att ttg 2208
Ala Cys Asn Ala Val Cys Ala Ile Ile His Asn Cys Gly Gly Ile Leu
      725                      730                      735

gca aag gac aat cct ctc gaa cac tct tgg aag aga tat aac tac cgc 2256
Ala Lys Asp Asn Pro Leu Glu His Ser Trp Lys Arg Tyr Asn Tyr Arg
      740                      745                      750

aat aga aat taa 2268
Asn Arg Asn
      755

```

&lt;210&gt; 32

&lt;211&gt; 755

&lt;212&gt; PRT

&lt;213&gt; Agrobacterium tumefaciens

&lt;400&gt; 32

```

Met Ser Ala Ser Pro Leu Leu Asp Asn Gln Cys Asp His Phe Ser Thr
  1                      5                      10                      15

Lys Met Val Asp Leu Ile Met Val Asp Lys Ala Asp Glu Leu Asp Arg
      20                      25                      30

Arg Val Ser Asp Ala Phe Ser Glu Arg Glu Ala Ser Arg Gly Arg Arg
      35                      40                      45

Ile Thr Gln Ile Ser Gly Glu Cys Ser Ala Gly Leu Ala Cys Lys Arg
      50                      55                      60

Leu Ala Asp Gly Arg Phe Pro Glu Ile Ser Thr Gly Glu Lys Val Ala
      65                      70                      75                      80

Ala Leu Ser Ala Tyr Ile Tyr Val Gly Lys Glu Ile Leu Gly Arg Ile
      85                      90                      95

Leu Glu Ser Glu Pro Trp Ala Arg Ala Arg Val Ser Gly Leu Val Ala
      100                      105                      110

Ile Asp Leu Ala Pro Phe Cys Met Asp Phe Ser Glu Ala Gln Leu Leu
      115                      120                      125

Gln Thr Leu Phe Leu Leu Ser Gly Lys Arg Cys Ala Ser Ser Asp Leu
      130                      135                      140

Ser His Phe Val Ala Ile Ser Ile Ser Lys Thr Ala Arg Ser Arg Thr
      145                      150                      155                      160

Leu Gln Met Pro Pro Tyr Glu Lys Gly Thr Thr Lys Arg Val Thr Gly
      165                      170                      175

Phe Thr Leu Thr Leu Glu Glu Ala Val Pro Phe Asp Met Val Ala Tyr
      180                      185                      190

Gly Arg Asn Leu Met Leu Lys Ala Ser Ala Gly Ser Phe Pro Thr Ile
      195                      200                      205

```



Asp 210	Leu	Tyr	Asp	Tyr	Arg	Ser	Phe	Phe	Asp	Gln	Cys	Ser	Asp	Ile	
Gly 225	Arg	Ile	Gly	Phe	Phe	Pro	Glu	Asp	Val	Pro	Lys	Pro	Lys	Val	Ala
Ile	Ile	Gly	Ala	Gly	Ile	Ser	Gly	Leu	Val	Val	Ala	Ser	Glu	Leu	Leu
His	Ala	Gly	Val	Asp	Asp	Val	Thr	Ile	Tyr	Glu	Ala	Ser	Asp	Arg	Val
Gly	Gly	Lys	Leu	Trp	Ser	His	Ala	Phe	Lys	Asp	Ala	Pro	Ser	Val	Val
Ala	Glu	Met	Gly	Ala	Met	Arg	Phe	Pro	Pro	Ala	Ala	Ser	Cys	Leu	Phe
Phe 305	Phe	Leu	Glu	Arg	Tyr	Gly	Leu	Ser	Ser	Met	Arg	Pro	Phe	Pro	Asn
Pro	Gly	Thr	Val	Asp	Thr	Asn	Leu	Val	Tyr	Gln	Gly	Leu	Arg	Tyr	Val
Trp	Lys	Ala	Gly	Gln	Gln	Pro	Pro	Lys	Leu	Phe	His	Arg	Val	Tyr	Ser
Gly	Trp	Arg	Ala	Phe	Leu	Arg	Asp	Gly	Phe	His	Glu	Gly	Asp	Ile	Val
Leu	Ala	Ser	Pro	Val	Val	Ile	Thr	Gln	Ala	Leu	Lys	Ser	Gly	Asp	Ile
Arg 385	Arg	Ala	His	Asp	Ser	Trp	Gln	Thr	Trp	Leu	Asn	Arg	Phe	Gly	Arg
Glu	Ser	Phe	Ser	Ser	Ala	Ile	Glu	Arg	Ile	Phe	Leu	Gly	Thr	His	Pro
Pro	Gly	Gly	Glu	Thr	Trp	Ser	Phe	Pro	His	Asp	Trp	Asp	Leu	Phe	Lys
Leu	Met	Gly	Ile	Gly	Ser	Gly	Gly	Phe	Gly	Pro	Val	Phe	Glu	Ser	Gly
Phe	Ile	Glu	Ile	Leu	Arg	Leu	Val	Ile	Asn	Gly	Tyr	Glu	Glu	Asn	Gln
Arg 465	Met	Cys	Ser	Glu	Gly	Ile	Ser	Glu	Leu	Pro	Arg	Arg	Ile	Ala	Ser
Gln	Val	Val	Asn	Gly	Val	Ser	Val	Ser	Gln	Arg	Ile	Arg	His	Val	Gln
Val	Arg	Ala	Ile	Glu	Lys	Glu	Lys	Thr	Lys	Ile	Lys	Ile	Arg	Leu	Lys
Ser	Gly	Ile	Ser	Glu	Leu	Tyr	Asp	Lys	Val	Val	Val	Thr	Ser	Gly	Leu
Ala	Asn	Ile	Gln	Leu	Arg	His	Cys	Leu	Thr	Cys	Asp	Thr	Thr	Ile	Phe
Arg 545	Ala	Pro	Val	Asn	Gln	Ala	Val	Asp	Asn	Ser	His	Met	Thr	Gly	Ser
Ser	Lys	Leu	Phe	Leu	Leu	Thr	Glu	Arg	Lys	Phe	Trp	Leu	Asp	His	Ile
Leu	Pro	Ser	Cys	Val	Leu	Met	Asp	Gly	Ile	Ala	Lys	Ala	Val	Tyr	Cys

Leu Asp Tyr Glu Pro Gln Asp Pro Asn Gly Lys Gly Leu Val Pro Pro  
 595 600 605  
 Thr Tyr Thr Trp Glu Asp Asp Ser His Lys Leu Leu Ala Val Pro Asp  
 610 615 620  
 Lys Lys Glu Arg Phe Cys Leu Leu Arg Asp Ala Ile Ser Arg Ser Phe  
 625 630 635 640  
 Pro Ala Phe Ala Gln His Leu Val Pro Ala Cys Ala Asp Tyr Asp Gln  
 645 650 655  
 Asn Val Val Gln His Asp Trp Leu Thr Asp Glu Asn Ala Gly Gly Ala  
 660 665 670  
 Phe Lys Leu Asn Arg Arg Gly Glu Asp Phe Tyr Ser Glu Glu Leu Phe  
 675 680 685  
 Phe Gln Ala Leu Asp Met Pro Asn Asp Thr Gly Val Tyr Leu Ala Gly  
 690 695 700  
 Cys Ser Cys Ser Phe Thr Gly Gly Trp Val Glu Gly Ala Ile Gln Thr  
 705 710 715 720  
 Ala Cys Asn Ala Val Cys Ala Ile Ile His Asn Cys Gly Gly Ile Leu  
 725 730 735  
 Ala Lys Asp Asn Pro Leu Glu His Ser Trp Lys Arg Tyr Asn Tyr Arg  
 740 745 750  
 Asn Arg Asn  
 755

&lt;210&gt; 33

&lt;211&gt; 1404

&lt;212&gt; DNA

&lt;213&gt; Agrobacterium tumefaciens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1401)

&lt;223&gt; coding for indole acetamide hydrolase

&lt;400&gt; 33

atg gtg ccc att acc tcg tta gca caa acc cta gaa cgc ctg aga cgg 48  
 Met Val Pro Ile Thr Ser Leu Ala Gln Thr Leu Glu Arg Leu Arg Arg  
 1 5 10 15  
 aaa gac tac tcc tgc tta gaa cta gta gaa act ctg ata gcg cgt tgc 96  
 Lys Asp Tyr Ser Cys Leu Glu Leu Val Glu Thr Leu Ile Ala Arg Cys  
 20 25 30  
 caa gct gca aaa cca tta aat gcc ctt ctg gct aca gac tgg gat ggc 144  
 Gln Ala Ala Lys Pro Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Gly  
 35 40 45  
 ttg cgg cga agc gcc aaa aaa aat gat cgt cat gga aac gcc gga tta 192  
 Leu Arg Arg Ser Ala Lys Lys Asn Asp Arg His Gly Asn Ala Gly Leu  
 50 55 60  
 ggt ctt tgc ggc att cca ctc tgt ttt aag gcg aac atc gcg acc ggc 240  
 Gly Leu Cys Gly Ile Pro Leu Cys Phe Lys Ala Asn Ile Ala Thr Gly  
 65 70 75 80  
 gta ttt cct aca agc gct gct act ccg gcg ctg ata aac cac ttg cca 288  
 Val Phe Pro Thr Ser Ala Ala Thr Pro Ala Leu Ile Asn His Leu Pro  
 85 90 95

aag	ata	cca	tcc	cgc	gtc	gca	gaa	aga	ctt	ttt	tca	gct	gga	gca	ctg	336
Lys	Ile	Pro	Ser	Arg	Val	Ala	Glu	Arg	Leu	Phe	Ser	Ala	Gly	Ala	Leu	
			100					105					110			
ccg	ggt	gcc	tcg	gga	aac	atg	cat	gag	tta	tcg	ttt	gga	att	acg	agc	384
Pro	Gly	Ala	Ser	Gly	Asn	Met	His	Glu	Leu	Ser	Phe	Gly	Ile	Thr	Ser	
		115					120					125				
aac	aac	tat	gcc	acc	ggt	gcg	gtg	cgg	aac	ccg	tgg	aat	cca	agt	ctg	432
Asn	Asn	Tyr	Ala	Thr	Gly	Ala	Val	Arg	Asn	Pro	Trp	Asn	Pro	Ser	Leu	
		130				135					140					
ata	cca	ggg	ggt	tca	agc	ggt	ggt	gtg	gct	gct	gcg	gtg	gca	agc	cga	480
Ile	Pro	Gly	Gly	Ser	Ser	Gly	Gly	Val	Ala	Ala	Ala	Val	Ala	Ser	Arg	
145						150				155					160	
ttg	atg	tta	ggc	ggc	ata	ggc	acg	gat	acc	ggt	gca	tct	ggt	cgc	cta	528
Leu	Met	Leu	Gly	Gly	Ile	Gly	Thr	Asp	Thr	Gly	Ala	Ser	Val	Arg	Leu	
			165					170						175		
ccg	gca	gcc	ctg	tgt	ggc	gta	gta	gga	ttt	cga	ccg	acg	ctt	ggt	cga	576
Pro	Ala	Ala	Leu	Cys	Gly	Val	Val	Gly	Phe	Arg	Pro	Thr	Leu	Gly	Arg	
			180					185					190			
tat	cca	aga	gat	cgg	ata	ata	ccg	ttc	agc	ccc	acc	cgg	gac	acc	gcc	624
Tyr	Pro	Arg	Asp	Arg	Ile	Ile	Pro	Phe	Ser	Pro	Thr	Arg	Asp	Thr	Ala	
		195					200					205				
gga	atc	ata	gcg	cag	tgc	gta	gcc	gat	gtt	ata	atc	ctc	gac	cag	gtg	672
Gly	Ile	Ile	Ala	Gln	Cys	Val	Ala	Asp	Val	Ile	Ile	Leu	Asp	Gln	Val	
	210					215					220					
att	tcc	gga	cgg	tcg	gcg	aaa	att	tca	ccc	atg	ccg	ctg	aag	ggg	ctt	720
Ile	Ser	Gly	Arg	Ser	Ala	Lys	Ile	Ser	Pro	Met	Pro	Leu	Lys	Gly	Leu	
225					230					235				240		
cgg	atc	ggc	ctc	ccc	act	acc	tac	ttt	tac	gat	gac	ctt	gat	gct	gat	768
Arg	Ile	Gly	Leu	Pro	Thr	Thr	Tyr	Phe	Tyr	Asp	Asp	Leu	Asp	Ala	Asp	
			245						250					255		
gtg	gcc	ttc	gca	gct	gaa	acg	acg	att	cgc	ttg	cta	gcc	aac	aga	ggc	816
Val	Ala	Phe	Ala	Ala	Glu	Thr	Thr	Ile	Arg	Leu	Leu	Ala	Asn	Arg	Gly	
		260						265					270			
gta	acc	ttt	ggt	gaa	gcc	gac	atc	ccc	cac	cta	gag	gaa	ttg	aac	agt	864
Val	Thr	Phe	Val	Glu	Ala	Asp	Ile	Pro	His	Leu	Glu	Glu	Leu	Asn	Ser	
		275				280						285				
ggg	gca	agt	ttg	cca	att	gcg	ctt	tac	gaa	ttt	cca	cac	gct	cta	aaa	912
Gly	Ala	Ser	Leu	Pro	Ile	Ala	Leu	Tyr	Glu	Phe	Pro	His	Ala	Leu	Lys	
	290					295					300					
aag	tat	ctc	gac	gat	ttt	gtg	gga	aca	gtt	tct	ttt	tct	gac	gtt	atc	960
Lys	Tyr	Leu	Asp	Asp	Phe	Val	Gly	Thr	Val	Ser	Phe	Ser	Asp	Val	Ile	
305					310				315					320		
aaa	gga	att	cgt	agc	ccc	gat	gta	gcg	aac	att	gtc	agt	gcg	caa	att	1008
Lys	Gly	Ile	Arg	Ser	Pro	Asp	Val	Ala	Asn	Ile	Val	Ser	Ala	Gln	Ile	
			325						330					335		
gat	ggg	cat	caa	att	tcc	aac	gat	gaa	tat	gaa	ctg	gcg	cgt	caa	tcc	1056
Asp	Gly	His	Gln	Ile	Ser	Asn	Asp	Glu	Tyr	Glu	Leu	Ala	Arg	Gln	Ser	
			340					345					350			
ttc	agg	cca	agg	ctc	cag	gcc	act	tat	cgg	aat	tac	ttc	aga	ctc	tat	1104
Phe	Arg	Pro	Arg	Leu	Gln	Ala	Thr	Tyr	Arg	Asn	Tyr	Phe	Arg	Leu	Tyr	
		355					360					365				

cag tta gat gca atc ctt ttc cca act gca ccc tta gcg gcc aaa gcc 1152  
 Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Ala Ala Lys Ala  
 370 375 380

ata ggt cag gag tcg tca gtc atc cac aat ggc tca atg atg aac act 1200  
 Ile Gly Gln Glu Ser Ser Val Ile His Asn Gly Ser Met Met Asn Thr  
 385 390 395 400

ttc aag atc tac gtg cga aat gtg gac cca agc agc aac gca ggc cta 1248  
 Phe Lys Ile Tyr Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu  
 405 410 415

cct ggg ttg agc ctt cct gcc tgc ctt aca cct gat cgc ttg cct gtt 1296  
 Pro Gly Leu Ser Leu Pro Ala Cys Leu Thr Pro Asp Arg Leu Pro Val  
 420 425 430

gga atg gaa att gat gga tta gcg ggg tca gac cac cgt ctg tta gca 1344  
 Gly Met Glu Ile Asp Gly Leu Ala Gly Ser Asp His Arg Leu Leu Ala  
 435 440 445

atc ggg gca gca tta gaa aaa gct ata aat ttt tct tcc ttt ccc gat 1392  
 Ile Gly Ala Ala Leu Glu Lys Ala Ile Asn Phe Ser Ser Phe Pro Asp  
 450 455 460

gct ttt aat tag 1404  
 Ala Phe Asn  
 465

<210> 34  
 <211> 467  
 <212> PRT  
 <213> Agrobacterium tumefaciens  
 <400> 34

Met Val Pro Ile Thr Ser Leu Ala Gln Thr Leu Glu Arg Leu Arg Arg  
 1 5 10 15

Lys Asp Tyr Ser Cys Leu Glu Leu Val Glu Thr Leu Ile Ala Arg Cys  
 20 25 30

Gln Ala Ala Lys Pro Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Gly  
 35 40 45

Leu Arg Arg Ser Ala Lys Lys Asn Asp Arg His Gly Asn Ala Gly Leu  
 50 55 60

Gly Leu Cys Gly Ile Pro Leu Cys Phe Lys Ala Asn Ile Ala Thr Gly  
 65 70 75 80

Val Phe Pro Thr Ser Ala Ala Thr Pro Ala Leu Ile Asn His Leu Pro  
 85 90 95

Lys Ile Pro Ser Arg Val Ala Glu Arg Leu Phe Ser Ala Gly Ala Leu  
 100 105 110

Pro Gly Ala Ser Gly Asn Met His Glu Leu Ser Phe Gly Ile Thr Ser  
 115 120 125

Asn Asn Tyr Ala Thr Gly Ala Val Arg Asn Pro Trp Asn Pro Ser Leu  
 130 135 140

Ile Pro Gly Gly Ser Ser Gly Gly Val Ala Ala Ala Val Ala Ser Arg  
 145 150 155 160

Leu Met Leu Gly Gly Ile Gly Thr Asp Thr Gly Ala Ser Val Arg Leu  
 165 170 175

Pro Ala Ala Leu Cys Gly Val Val Gly Phe Arg Pro Thr Leu Gly Arg  
 180 185 190

Tyr Pro Arg Asp Arg Ile Ile Pro Phe Ser Pro Thr Arg Asp Thr Ala  
 195 200 205  
 Gly Ile Ile Ala Gln Cys Val Ala Asp Val Ile Ile Leu Asp Gln Val  
 210 215 220  
 Ile Ser Gly Arg Ser Ala Lys Ile Ser Pro Met Pro Leu Lys Gly Leu  
 225 230 235 240  
 Arg Ile Gly Leu Pro Thr Thr Tyr Phe Tyr Asp Asp Leu Asp Ala Asp  
 245 250 255  
 Val Ala Phe Ala Ala Glu Thr Thr Ile Arg Leu Leu Ala Asn Arg Gly  
 260 265 270  
 Val Thr Phe Val Glu Ala Asp Ile Pro His Leu Glu Glu Leu Asn Ser  
 275 280 285  
 Gly Ala Ser Leu Pro Ile Ala Leu Tyr Glu Phe Pro His Ala Leu Lys  
 290 295 300  
 Lys Tyr Leu Asp Asp Phe Val Gly Thr Val Ser Phe Ser Asp Val Ile  
 305 310 315 320  
 Lys Gly Ile Arg Ser Pro Asp Val Ala Asn Ile Val Ser Ala Gln Ile  
 325 330 335  
 Asp Gly His Gln Ile Ser Asn Asp Glu Tyr Glu Leu Ala Arg Gln Ser  
 340 345 350  
 Phe Arg Pro Arg Leu Gln Ala Thr Tyr Arg Asn Tyr Phe Arg Leu Tyr  
 355 360 365  
 Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Ala Ala Lys Ala  
 370 375 380  
 Ile Gly Gln Glu Ser Ser Val Ile His Asn Gly Ser Met Met Asn Thr  
 385 390 395 400  
 Phe Lys Ile Tyr Val Arg Asn Val Asp Pro Ser Ser Asn Ala Gly Leu  
 405 410 415  
 Pro Gly Leu Ser Leu Pro Ala Cys Leu Thr Pro Asp Arg Leu Pro Val  
 420 425 430  
 Gly Met Glu Ile Asp Gly Leu Ala Gly Ser Asp His Arg Leu Leu Ala  
 435 440 445  
 Ile Gly Ala Ala Leu Glu Lys Ala Ile Asn Phe Ser Ser Phe Pro Asp  
 450 455 460  
 Ala Phe Asn  
 465

&lt;210&gt; 35

&lt;211&gt; 1419

&lt;212&gt; DNA

&lt;213&gt; Agrobacterium vitis

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1416)

&lt;223&gt; coding for indole acetamide hydrolase

&lt;400&gt; 35

atg gtg acc cta ggt tca atc aag gaa acc ctg gaa tgt ctc agg ctg

Met Val Thr Leu Gly Ser Ile Lys Glu Thr Leu Glu Cys Leu Arg Leu

1

5

10

15

48

aaa aaa tac tcc tgt tcc gaa ctg gct gaa acc ata ata gcc cgt tgc	96
Lys Lys Tyr Ser Cys Ser Glu Leu Ala Glu Thr Ile Ile Ala Arg Cys	
20 25 30	
gaa gcc gcg aaa tct ctc aat gct ctt ctg gcg act gac tgg gat tac	144
Glu Ala Ala Lys Ser Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Tyr	
35 40 45	
ctg cgg cgt aat gcc aag aaa gta gat gaa gat gga agc gcc ggc gag	192
Leu Arg Arg Asn Ala Lys Lys Val Asp Glu Asp Gly Ser Ala Gly Glu	
50 55 60	
ggt ctt gcc ggc atc ccg ctg tgt tct aaa gcg aac att gca aca ggc	240
Gly Leu Ala Gly Ile Pro Leu Cys Ser Lys Ala Asn Ile Ala Thr Gly	
65 70 75 80	
ata ttc cca gca agc gcg gcc acg ccg gcg ctt gat gaa cat tta cct	288
Ile Phe Pro Ala Ser Ala Ala Thr Pro Ala Leu Asp Glu His Leu Pro	
85 90 95	
aca aca cca gcc ggc gtc cgt aaa ccg ctt cta gac gct ggg gca ctg	336
Thr Thr Pro Ala Gly Val Arg Lys Pro Leu Leu Asp Ala Gly Ala Leu	
100 105 110	
ata ggc gct tcg gga aac atg cat gag tta tcg ttt ggc att acc agt	384
Ile Gly Ala Ser Gly Asn Met His Glu Leu Ser Phe Gly Ile Thr Ser	
115 120 125	
aac aac cac gcc act ggt gcg gtg aga aac ccc tgg aat ccc agc tta	432
Asn Asn His Ala Thr Gly Ala Val Arg Asn Pro Trp Asn Pro Ser Leu	
130 135 140	
ata cca gga ggc tcg agc ggc ggc gtg gct gct gct gta gca tca cgg	480
Ile Pro Gly Gly Ser Ser Gly Gly Val Ala Ala Ala Val Ala Ser Arg	
145 150 155 160	
tta atg ctc ggc gga att ggc acc gac acg ggg gct tcg gtc cgc cta	528
Leu Met Leu Gly Gly Ile Gly Thr Asp Thr Gly Ala Ser Val Arg Leu	
165 170 175	
cct gca tcc cta tgt ggc gta gtg gga ttc cgc ccg acg atc ggc aga	576
Pro Ala Ser Leu Cys Gly Val Val Gly Phe Arg Pro Thr Ile Gly Arg	
180 185 190	
tat cct gga gac cga att gtg ccg gtt agc ccc acc cgc gat aca gcc	624
Tyr Pro Gly Asp Arg Ile Val Pro Val Ser Pro Thr Arg Asp Thr Ala	
195 200 205	
gga att atc gca cag agc gtt cct gat gtg ata ctc ctt gac caa atc	672
Gly Ile Ile Ala Gln Ser Val Pro Asp Val Ile Leu Leu Asp Gln Ile	
210 215 220	
att tgc ggg aag ctc acg acc cac caa cct gta ccc ctg gag gga tta	720
Ile Cys Gly Lys Leu Thr Thr His Gln Pro Val Pro Leu Glu Gly Leu	
225 230 235 240	
cgt atc ggc ttg cca acc act tac ttt tac gat gac ctt gat gct gat	768
Arg Ile Gly Leu Pro Thr Thr Tyr Phe Tyr Asp Asp Leu Asp Ala Asp	
245 250 255	
gtg gcc ttc gca gct gaa aac ctt atc acg ctg ctg gcc agc aag ggt	816
Val Ala Phe Ala Ala Glu Asn Leu Ile Thr Leu Leu Ala Ser Lys Gly	
260 265 270	
gta acc ttt gtt aag gcc gag att cca gat ctg cag cgt ctg aac atc	864
Val Thr Phe Val Lys Ala Glu Ile Pro Asp Leu Gln Arg Leu Asn Ile	
275 280 285	

ggg gtt agc ttt cct att gcc ctg tac gag ttt ccg ttc gcc cta caa 912  
 Gly Val Ser Phe Pro Ile Ala Leu Tyr Glu Phe Pro Phe Ala Leu Gln  
 290 295 300

aag tat atc gat gac ttt gtg aag gat gtg tct ttt tct gac gtc atc 960  
 Lys Tyr Ile Asp Asp Phe Val Lys Asp Val Ser Phe Ser Asp Val Ile  
 305 310 315 320

aaa gga att cgt agc cct gat gta gcc aac att gcc aat gct caa att 1008  
 Lys Gly Ile Arg Ser Pro Asp Val Ala Asn Ile Ala Asn Ala Gln Ile  
 325 330 335

gat gga cat caa att tcc aaa gct tca tat gaa ctg gcg cga caa tct 1056  
 Asp Gly His Gln Ile Ser Lys Ala Ser Tyr Glu Leu Ala Arg Gln Ser  
 340 345 350

ttc aga cca aag ctg caa gcc gcc tac cat gat tac ttc aag ctg cac 1104  
 Phe Arg Pro Lys Leu Gln Ala Ala Tyr His Asp Tyr Phe Lys Leu His  
 355 360 365

cag cta gac gcg atc ctt ttc ccg aca gct ccc ctg aca gcc aaa ccg 1152  
 Gln Leu Asp Ala Ile Leu Phe Pro Thr Ala Pro Leu Thr Ala Lys Pro  
 370 375 380

atc ggc caa gat tta tcg gtg atg cac aat ggc gta atg gcc gac acg 1200  
 Ile Gly Gln Asp Leu Ser Val Met His Asn Gly Val Met Ala Asp Thr  
 385 390 395 400

ttt aaa atc ttc gtg cga aat gtg gat ccg ggg agc aac gca ggc ctg 1248  
 Phe Lys Ile Phe Val Arg Asn Val Asp Pro Gly Ser Asn Ala Gly Leu  
 405 410 415

cca gga tta agc ctt ccc gtt tct ctt act tca aag ggt ttg cct att 1296  
 Pro Gly Leu Ser Leu Pro Val Ser Leu Thr Ser Lys Gly Leu Pro Ile  
 420 425 430

gga atg gaa atc gat gga tta gcg ggc atg gac gac cgt ttg cta gca 1344  
 Gly Met Glu Ile Asp Gly Leu Ala Gly Met Asp Asp Arg Leu Leu Ala  
 435 440 445

atc gga gcg gca cta gag gaa gcg ata gct ttt cat aat tta cct gac 1392  
 Ile Gly Ala Ala Leu Glu Glu Ala Ile Ala Phe His Asn Leu Pro Asp  
 450 455 460

ttc ccg aaa gtc gag aca aac tac tga 1419  
 Phe Pro Lys Val Glu Thr Asn Tyr  
 465 470

<210> 36  
 <211> 472  
 <212> PRT  
 <213> Agrobacterium vitis

<400> 36  
 Met Val Thr Leu Gly Ser Ile Lys Glu Thr Leu Glu Cys Leu Arg Leu  
 1 5 10 15  
 Lys Lys Tyr Ser Cys Ser Glu Leu Ala Glu Thr Ile Ile Ala Arg Cys  
 20 25 30  
 Glu Ala Ala Lys Ser Leu Asn Ala Leu Leu Ala Thr Asp Trp Asp Tyr  
 35 40 45  
 Leu Arg Arg Asn Ala Lys Lys Val Asp Glu Asp Gly Ser Ala Gly Glu  
 50 55 60  
 Gly Leu Ala Gly Ile Pro Leu Cys Ser Lys Ala Asn Ile Ala Thr Gly  
 65 70 75 80

Ile	Phe	Pro	Ala	Ser	Ala	Ala	Thr	Pro	Ala	Leu	Asp	Glu	His	Leu	Pro
				85					90					95	
Thr	Thr	Pro	Ala	Gly	Val	Arg	Lys	Pro	Leu	Leu	Asp	Ala	Gly	Ala	Leu
			100					105					110		
Ile	Gly	Ala	Ser	Gly	Asn	Met	His	Glu	Leu	Ser	Phe	Gly	Ile	Thr	Ser
		115					120					125			
Asn	Asn	His	Ala	Thr	Gly	Ala	Val	Arg	Asn	Pro	Trp	Asn	Pro	Ser	Leu
		130				135					140				
Ile	Pro	Gly	Gly	Ser	Ser	Gly	Gly	Val	Ala	Ala	Ala	Val	Ala	Ser	Arg
145					150					155					160
Leu	Met	Leu	Gly	Gly	Ile	Gly	Thr	Asp	Thr	Gly	Ala	Ser	Val	Arg	Leu
				165				170						175	
Pro	Ala	Ser	Leu	Cys	Gly	Val	Val	Gly	Phe	Arg	Pro	Thr	Ile	Gly	Arg
			180					185					190		
Tyr	Pro	Gly	Asp	Arg	Ile	Val	Pro	Val	Ser	Pro	Thr	Arg	Asp	Thr	Ala
		195					200					205			
Gly	Ile	Ile	Ala	Gln	Ser	Val	Pro	Asp	Val	Ile	Leu	Leu	Asp	Gln	Ile
	210					215					220				
Ile	Cys	Gly	Lys	Leu	Thr	Thr	His	Gln	Pro	Val	Pro	Leu	Glu	Gly	Leu
225					230					235					240
Arg	Ile	Gly	Leu	Pro	Thr	Thr	Tyr	Phe	Tyr	Asp	Asp	Leu	Asp	Ala	Asp
				245					250					255	
Val	Ala	Phe	Ala	Ala	Glu	Asn	Leu	Ile	Thr	Leu	Leu	Ala	Ser	Lys	Gly
			260					265					270		
Val	Thr	Phe	Val	Lys	Ala	Glu	Ile	Pro	Asp	Leu	Gln	Arg	Leu	Asn	Ile
		275					280					285			
Gly	Val	Ser	Phe	Pro	Ile	Ala	Leu	Tyr	Glu	Phe	Pro	Phe	Ala	Leu	Gln
	290					295					300				
Lys	Tyr	Ile	Asp	Asp	Phe	Val	Lys	Asp	Val	Ser	Phe	Ser	Asp	Val	Ile
305					310					315					320
Lys	Gly	Ile	Arg	Ser	Pro	Asp	Val	Ala	Asn	Ile	Ala	Asn	Ala	Gln	Ile
				325					330					335	
Asp	Gly	His	Gln	Ile	Ser	Lys	Ala	Ser	Tyr	Glu	Leu	Ala	Arg	Gln	Ser
			340					345					350		
Phe	Arg	Pro	Lys	Leu	Gln	Ala	Ala	Tyr	His	Asp	Tyr	Phe	Lys	Leu	His
		355					360					365			
Gln	Leu	Asp	Ala	Ile	Leu	Phe	Pro	Thr	Ala	Pro	Leu	Thr	Ala	Lys	Pro
	370					375					380				
Ile	Gly	Gln	Asp	Leu	Ser	Val	Met	His	Asn	Gly	Val	Met	Ala	Asp	Thr
385					390					395					400
Phe	Lys	Ile	Phe	Val	Arg	Asn	Val	Asp	Pro	Gly	Ser	Asn	Ala	Gly	Leu
				405					410					415	
Pro	Gly	Leu	Ser	Leu	Pro	Val	Ser	Leu	Thr	Ser	Lys	Gly	Leu	Pro	Ile
			420					425					430		
Gly															



Phe Pro Lys Val Glu Thr Asn Tyr  
465 470

<210> 37

<211> 1263

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (1)..(1260)

<223> coding for 5-methylthioribose kinase

<400> 37

atg tct ttt gag gag ttt acg ccg tta aac gag aag tct ctt gta gac	48
Met Ser Phe Glu Glu Phe Thr Pro Leu Asn Glu Lys Ser Leu Val Asp	
1 5 10 15	
tac atc aag tca aca cct gct ctc tct tcc aag atc gga gcc gac aag	96
Tyr Ile Lys Ser Thr Pro Ala Leu Ser Ser Lys Ile Gly Ala Asp Lys	
20 25 30	
tcc gat gat gat ttg gtt atc aaa gaa gtt gga gat ggc aat ctc aat	144
Ser Asp Asp Asp Leu Val Ile Lys Glu Val Gly Asp Gly Asn Leu Asn	
35 40 45	
ttc gtt ttc atc gtt gtt gga tcc tct ggt tct ctt gtc atc aaa cag	192
Phe Val Phe Ile Val Val Gly Ser Ser Gly Ser Leu Val Ile Lys Gln	
50 55 60	
gct ctt cca tat att cgc tgt atc ggt gaa tca tgg cca atg acg aaa	240
Ala Leu Pro Tyr Ile Arg Cys Ile Gly Glu Ser Trp Pro Met Thr Lys	
65 70 75 80	
gaa aga gct tat ttt gaa gca aca act ttg aga aag cat gga aat tta	288
Glu Arg Ala Tyr Phe Glu Ala Thr Thr Leu Arg Lys His Gly Asn Leu	
85 90 95	
tca cct gat cat gtt cct gaa gtc tac cat ttt gac aga aca atg gcg	336
Ser Pro Asp His Val Pro Glu Val Tyr His Phe Asp Arg Thr Met Ala	
100 105 110	
ttg att gga atg aga tac ctt gag cct cct cat atc att ctc cgc aaa	384
Leu Ile Gly Met Arg Tyr Leu Glu Pro Pro His Ile Ile Leu Arg Lys	
115 120 125	
gga ctc att gct ggg att gag tat cct ttc ctc gca gac cac atg tct	432
Gly Leu Ile Ala Gly Ile Glu Tyr Pro Phe Leu Ala Asp His Met Ser	
130 135 140	
gat tac atg gcg aag act ctc ttc ttc act tct ctc ctc tat cac gat	480
Asp Tyr Met Ala Lys Thr Leu Phe Phe Thr Ser Leu Leu Tyr His Asp	
145 150 155 160	
acc aca gag cac aga aga gca gta acc gaa ttt tgt ggt aat gtg gag	528
Thr Thr Glu His Arg Arg Ala Val Thr Glu Phe Cys Gly Asn Val Glu	
165 170 175	
tta tgc cga tta acg gag caa gtt gtg ttt tcg gac cca tat aga gtt	576
Leu Cys Arg Leu Thr Glu Gln Val Val Phe Ser Asp Pro Tyr Arg Val	
180 185 190	
tcc aca ttt aat cgt tgg act tca cct tat ctt gat gat gat gct aag	624
Ser Thr Phe Asn Arg Trp Thr Ser Pro Tyr Leu Asp Asp Asp Ala Lys	
195 200 205	

```

gct gtg cgc gaa gac agt gcc ttg aag ctc gaa atc gca gag cta aaa 672
Ala Val Arg Glu Asp Ser Ala Leu Lys Leu Glu Ile Ala Glu Leu Lys
210 215 220

tcg atg ttc tgt gaa aga gct caa gct tta ata cat ggt gat ctt cat 720
Ser Met Phe Cys Glu Arg Ala Gln Ala Leu Ile His Gly Asp Leu His
225 230 235 240

act ggt tct gtc atg gtt act caa gat tca acg caa gtt ata gat cca 768
Thr Gly Ser Val Met Val Thr Gln Asp Ser Thr Gln Val Ile Asp Pro
245 250 255

gag ttt tcg ttc tat gga ccg atg ggt ttc gat att ggc gct tat ctt 816
Glu Phe Ser Phe Tyr Gly Pro Met Gly Phe Asp Ile Gly Ala Tyr Leu
260 265 270

ggt aac ttg ata cta gct ttc ttt gca caa gat gga cac gcc act cag 864
Gly Asn Leu Ile Leu Ala Phe Phe Ala Gln Asp Gly His Ala Thr Gln
275 280 285

gaa aat gat cga aaa gaa tac aag cag tgg atc ttg aga acc att gag 912
Glu Asn Asp Arg Lys Glu Tyr Lys Gln Trp Ile Leu Arg Thr Ile Glu
290 295 300

caa act tgg aat ttg ttt aac aaa agg ttc att gcg cta tgg gat caa 960
Gln Thr Trp Asn Leu Phe Asn Lys Arg Phe Ile Ala Leu Trp Asp Gln
305 310 315 320

aac aaa gat gga cca ggc gaa gca tac ctt gca gat atc tat aac aat 1008
Asn Lys Asp Gly Pro Gly Glu Ala Tyr Leu Ala Asp Ile Tyr Asn Asn
325 330 335

acc gag gtt ttg aag ttt gtt caa gaa aac tac atg agg aat ttg ttg 1056
Thr Glu Val Leu Lys Phe Val Gln Glu Asn Tyr Met Arg Asn Leu Leu
340 345 350

cat gac tca ctc gga ttc ggc gct gca aag atg att agg aga att gtg 1104
His Asp Ser Leu Gly Phe Gly Ala Ala Lys Met Ile Arg Arg Ile Val
355 360 365

gga gtg gca cat gtt gag gac ttt gaa tca atc gaa gaa gat aag cga 1152
Gly Val Ala His Val Glu Asp Phe Glu Ser Ile Glu Glu Asp Lys Arg
370 375 380

aga gct att tgc gag aga agt gca ctc gag ttt gcg aag atg ctt ctc 1200
Arg Ala Ile Cys Glu Arg Ser Ala Leu Glu Phe Ala Lys Met Leu Leu
385 390 395 400

aag gaa agg aga aag ttt aag agt atc ggt gaa gtt gtt tca gca att 1248
Lys Glu Arg Arg Lys Phe Lys Ser Ile Gly Glu Val Val Ser Ala Ile
405 410 415

caa caa caa agc taa 1263
Gln Gln Gln Ser
420

<210> 38
<211> 420
<212> PRT
<213> Arabidopsis thaliana

<400> 38
Met Ser Phe Glu Glu Phe Thr Pro Leu Asn Glu Lys Ser Leu Val Asp
1 5 10 15
Tyr Ile Lys Ser Thr Pro Ala Leu Ser Ser Lys Ile Gly Ala Asp Lys
20 25 30

```

Ser Asp Asp Asp Leu Val Ile Lys Glu Val Gly Asp Gly Asn Leu Asn  
                   35                                  40                                  45  
 Phe Val Phe Ile Val Val Gly Ser Ser Gly Ser Leu Val Ile Lys Gln  
           50                                  55                                  60  
 Ala Leu Pro Tyr Ile Arg Cys Ile Gly Glu Ser Trp Pro Met Thr Lys  
           65                                  70                                  75                                  80  
 Glu Arg Ala Tyr Phe Glu Ala Thr Thr Leu Arg Lys His Gly Asn Leu  
                                   85                                  90                                  95  
 Ser Pro Asp His Val Pro Glu Val Tyr His Phe Asp Arg Thr Met Ala  
                   100                                  105                                  110  
 Leu Ile Gly Met Arg Tyr Leu Glu Pro Pro His Ile Ile Leu Arg Lys  
           115                                  120                                  125  
 Gly Leu Ile Ala Gly Ile Glu Tyr Pro Phe Leu Ala Asp His Met Ser  
           130                                  135                                  140  
 Asp Tyr Met Ala Lys Thr Leu Phe Phe Thr Ser Leu Leu Tyr His Asp  
   145                                  150                                  155                                  160  
 Thr Thr Glu His Arg Arg Ala Val Thr Glu Phe Cys Gly Asn Val Glu  
                                   165                                  170                                  175  
 Leu Cys Arg Leu Thr Glu Gln Val Val Phe Ser Asp Pro Tyr Arg Val  
                   180                                  185                                  190  
 Ser Thr Phe Asn Arg Trp Thr Ser Pro Tyr Leu Asp Asp Asp Ala Lys  
           195                                  200                                  205  
 Ala Val Arg Glu Asp Ser Ala Leu Lys Leu Glu Ile Ala Glu Leu Lys  
           210                                  215                                  220  
 Ser Met Phe Cys Glu Arg Ala Gln Ala Leu Ile His Gly Asp Leu His  
   225                                  230                                  235                                  240  
 Thr Gly Ser Val Met Val Thr Gln Asp Ser Thr Gln Val Ile Asp Pro  
                                   245                                  250                                  255  
 Glu Phe Ser Phe Tyr Gly Pro Met Gly Phe Asp Ile Gly Ala Tyr Leu  
                   260                                  265                                  270  
 Gly Asn Leu Ile Leu Ala Phe Phe Ala Gln Asp Gly His Ala Thr Gln  
           275                                  280                                  285  
 Glu Asn Asp Arg Lys Glu Tyr Lys Gln Trp Ile Leu Arg Thr Ile Glu  
           290                                  295                                  300  
 Gln Thr Trp Asn Leu Phe Asn Lys Arg Phe Ile Ala Leu Trp Asp Gln  
   305                                  310                                  315                                  320  
 Asn Lys Asp Gly Pro Gly Glu Ala Tyr Leu Ala Asp Ile Tyr Asn Asn  
                                   325                                  330                                  335  
 Thr Glu Val Leu Lys Phe Val Gln Glu Asn Tyr Met Arg Asn Leu Leu  
                   340                                  345                                  350  
 His Asp Ser Leu Gly Phe Gly Ala Ala Lys Met Ile Arg Arg Ile Val  
           355                                  360                                  365  
 Gly Val Ala His Val Glu Asp Phe Glu Ser Ile Glu Glu Asp Lys Arg  
           370                                  375                                  380  
 Arg Ala Ile Cys Glu Arg Ser Ala Leu Glu Phe Ala Lys Met Leu Leu  
   385                                  390                                  395                                  400  
 Lys Glu Arg Arg Lys Phe Lys Ser Ile Gly Glu Val Val Ser Ala Ile  
                                   405                                  410                                  415

Gln Gln Gln Ser  
420

<210> 39  
<211> 1200  
<212> DNA  
<213> *Klebsiella pneumoniae*  
<220>  
<221> CDS  
<222> (1)..(1197)  
<223> coding for 5-methylthioribose kinase

```

<400> 39
atg tcg caa tac cat acc ttc acc gcc cac gat gcc gtg gct tac gcg 48
Met Ser Gln Tyr His Thr Phe Thr Ala His Asp Ala Val Ala Tyr Ala
  1             5             10             15

caa cag ttc gcc ggc atc gac aac cca tct gag ctg gtc agc gcg cag 96
Gln Gln Phe Ala Gly Ile Asp Asn Pro Ser Glu Leu Val Ser Ala Gln
             20             25             30

gaa gtg ggc gat ggc aac ctc aat ctg gtg ttt aaa gtg ttc gat cgt 144
Glu Val Gly Asp Gly Asn Leu Asn Leu Val Phe Lys Val Phe Asp Arg
             35             40             45

cag ggc gtc agc cgg gcg atc gtc aaa cag gcc ctg ccc tac gtg cgc 192
Gln Gly Val Ser Arg Ala Ile Val Lys Gln Ala Leu Pro Tyr Val Arg
             50             55             60

tgc gtc ggc gaa tcc tgg ccg ctg acc ctc gac cgc gcc cgt ctc gaa 240
Cys Val Gly Glu Ser Trp Pro Leu Thr Leu Asp Arg Ala Arg Leu Glu
  65             70             75             80

gcg cag acc ctg gtc gcc cac tat cag cac agc ccg cag cac acg gta 288
Ala Gln Thr Leu Val Ala His Tyr Gln His Ser Pro Gln His Thr Val
             85             90             95

aaa atc cat cac ttt gat ccc gag ctg gcg gtg atg gtg atg gaa gat 336
Lys Ile His His Phe Asp Pro Glu Leu Ala Val Met Val Met Glu Asp
             100            105            110

ctt tcc gac cac cgc atc tgg cgc gga gag ctt atc gct aac gtc tac 384
Leu Ser Asp His Arg Ile Trp Arg Gly Glu Leu Ile Ala Asn Val Tyr
             115            120            125

tat ccc cag gcg gcc cgc cag ctt ggc gac tat ctg gcg cag gtg ttg 432
Tyr Pro Gln Ala Ala Arg Gln Leu Gly Asp Tyr Leu Ala Gln Val Leu
             130            135            140

ttc cac acc agc gat ttc tac ctc cat ccc cac gag aaa aag gcg cag 480
Phe His Thr Ser Asp Phe Tyr Leu His Pro His Glu Lys Lys Ala Gln
             145            150            155            160

gtg gcg cag ttt att aac ccg gcg atg tgc gag atc acc gag gat ctg 528
Val Ala Gln Phe Ile Asn Pro Ala Met Cys Glu Ile Thr Glu Asp Leu
             165            170            175

ttc ttt aac gac ccg tat cag atc cac gag cgc aat aac tac ccg gcg 576
Phe Phe Asn Asp Pro Tyr Gln Ile His Glu Arg Asn Asn Tyr Pro Ala
             180            185            190

gag ctg gag gcc gat gtc gcc gcc ctg cgc gac gac gcc cag ctt aag 624
Glu Leu Glu Ala Asp Val Ala Ala Leu Arg Asp Asp Ala Gln Leu Lys
             195            200            205

```

```

ctg gcg gtg gcg gcg ctg aag cac cgt ttc ttt gcc cat gcg gaa gcg 672
Leu Ala Val Ala Ala Leu Lys His Arg Phe Phe Ala His Ala Glu Ala
210 215 220

ctg ctg cac ggc gat atc cac agc ggg tcg atc ttc gtt gcc gaa ggt 720
Leu Leu His Gly Asp Ile His Ser Gly Ser Ile Phe Val Ala Glu Gly
225 230 235 240

agc ctg aag gcc atc gac gcc gag ttc ggc tac ttc ggc ccc atc ggc 768
Ser Leu Lys Ala Ile Asp Ala Glu Phe Gly Tyr Phe Gly Pro Ile Gly
245 250 255

ttc gat atc ggc acc gcc atc ggc aac ctg ctg ctg aac tac tgc ggc 816
Phe Asp Ile Gly Thr Ala Ile Gly Asn Leu Leu Leu Asn Tyr Cys Gly
260 265 270

ctg ccg ggc cag ctc ggc att cgc gat gcc gcc gcc gcg cgc gag cag 864
Leu Pro Gly Gln Leu Gly Ile Arg Asp Ala Ala Ala Ala Arg Glu Gln
275 280 285

cgg ctg aac gac atc cac cag ctg tgg acc acc ttc gcc gag cgc ttc 912
Arg Leu Asn Asp Ile His Gln Leu Trp Thr Thr Phe Ala Glu Arg Phe
290 295 300

cag gcg ctg gcg gcg gag aaa acc cgc gac gcg gcg ctg gct tac ccc 960
Gln Ala Leu Ala Ala Glu Lys Thr Arg Asp Ala Ala Leu Ala Tyr Pro
305 310 315 320

ggc tac gcc tcc gcc ttt ctg aag aaa gtc tgg gcg gac gcg gtc ggc 1008
Gly Tyr Ala Ser Ala Phe Leu Lys Lys Val Trp Ala Asp Ala Val Gly
325 330 335

ttc tgc ggc agc gaa ctg atc cgc cgc agc gtc gga ctg tcg cac gtc 1056
Phe Cys Gly Ser Glu Leu Ile Arg Arg Ser Val Gly Leu Ser His Val
340 345 350

gcg gat atc gac act atc cag gac gac gcc atg cgt cat gag tgc ctg 1104
Ala Asp Ile Asp Thr Ile Gln Asp Asp Ala Met Arg His Glu Cys Leu
355 360 365

cgc cac gcc att acc ctg ggc aga gcg ctg atc gtg ctg gcc gag cgt 1152
Arg His Ala Ile Thr Leu Gly Arg Ala Leu Ile Val Leu Ala Glu Arg
370 375 380

atc gac agc gtc gac gag ctg ctg gcg cgg gta cgc cag tac agc tga 1200
Ile Asp Ser Val Asp Glu Leu Leu Ala Arg Val Arg Gln Tyr Ser
385 390 395

```

&lt;210&gt; 40

&lt;211&gt; 399

&lt;212&gt; PRT

<213> *Klebsiella pneumoniae*

&lt;400&gt; 40

```

Met Ser Gln Tyr His Thr Phe Thr Ala His Asp Ala Val Ala Tyr Ala
1 5 10 15

Gln Gln Phe Ala Gly Ile Asp Asn Pro Ser Glu Leu Val Ser Ala Gln
20 25 30

Glu Val Gly Asp Gly Asn Leu Asn Leu Val Phe Lys Val Phe Asp Arg
35 40 45

Gln Gly Val Ser Arg Ala Ile Val Lys Gln Ala Leu Pro Tyr Val Arg
50 55 60

Cys Val Gly Glu Ser Trp Pro Leu Thr Leu Asp Arg Ala Arg Leu Glu
65 70 75 80

```

Ala	Gln	Thr	Leu	Val	Ala	His	Tyr	Gln	His	Ser	Pro	Gln	His	Thr	Val	85	90	95
Lys	Ile	His	His	Phe	Asp	Pro	Glu	Leu	Ala	Val	Met	Val	Met	Glu	Asp	100	105	110
Leu	Ser	Asp	His	Arg	Ile	Trp	Arg	Gly	Glu	Leu	Ile	Ala	Asn	Val	Tyr	115	120	125
Tyr	Pro	Gln	Ala	Ala	Arg	Gln	Leu	Gly	Asp	Tyr	Leu	Ala	Gln	Val	Leu	130	135	140
Phe	His	Thr	Ser	Asp	Phe	Tyr	Leu	His	Pro	His	Glu	Lys	Lys	Ala	Gln	145	150	155
Val	Ala	Gln	Phe	Ile	Asn	Pro	Ala	Met	Cys	Glu	Ile	Thr	Glu	Asp	Leu	165	170	175
Phe	Phe	Asn	Asp	Pro	Tyr	Gln	Ile	His	Glu	Arg	Asn	Asn	Tyr	Pro	Ala	180	185	190
Glu	Leu	Glu	Ala	Asp	Val	Ala	Ala	Leu	Arg	Asp	Asp	Ala	Gln	Leu	Lys	195	200	205
Leu	Ala	Val	Ala	Ala	Leu	Lys	His	Arg	Phe	Phe	Ala	His	Ala	Glu	Ala	210	215	220
Leu	Leu	His	Gly	Asp	Ile	His	Ser	Gly	Ser	Ile	Phe	Val	Ala	Glu	Gly	225	230	235
Ser	Leu	Lys	Ala	Ile	Asp	Ala	Glu	Phe	Gly	Tyr	Phe	Gly	Pro	Ile	Gly	245	250	255
Phe	Asp	Ile	Gly	Thr	Ala	Ile	Gly	Asn	Leu	Leu	Leu	Asn	Tyr	Cys	Gly	260	265	270
Leu	Pro	Gly	Gln	Leu	Gly	Ile	Arg	Asp	Ala	Ala	Ala	Ala	Arg	Glu	Gln	275	280	285
Arg	Leu	Asn	Asp	Ile	His	Gln	Leu	Trp	Thr	Thr	Phe	Ala	Glu	Arg	Phe	290	295	300
Gln	Ala	Leu	Ala	Ala	Glu	Lys	Thr	Arg	Asp	Ala	Ala	Leu	Ala	Tyr	Pro	305	310	315
Gly	Tyr	Ala	Ser	Ala	Phe	Leu	Lys	Lys	Val	Trp	Ala	Asp	Ala	Val	Gly	325	330	335
Phe	Cys	Gly	Ser	Glu	Leu	Ile	Arg	Arg	Ser	Val	Gly	Leu	Ser	His	Val	340	345	350
Ala	Asp	Ile	Asp	Thr	Ile	Gln	Asp	Asp	Ala	Met	Arg	His	Glu	Cys	Leu	355	360	365
Arg	His	Ala	Ile	Thr	Leu	Gly	Arg	Ala	Leu	Ile	Val	Leu	Ala	Glu	Arg	370	375	380
Ile	Asp	Ser	Val	Asp	Glu	Leu	Leu	Ala	Arg	Val	Arg	Gln	Tyr	Ser		385	390	395

&lt;210&gt; 41

&lt;211&gt; 1140

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1137)

&lt;223&gt; coding for alcohol dehydrogenase

&lt;400&gt; 41

atg tct acc acc gga cag att att cga tgc aaa gct gct gtg gca tgg	48
Met Ser Thr Thr Gly Gln Ile Ile Arg Cys Lys Ala Ala Val Ala Trp	
1 5 10 15	
gaa gcc gga aag cca ctg gtg atc gag gaa gtg gag gtt gct cca ccg	96
Glu Ala Gly Lys Pro Leu Val Ile Glu Glu Val Glu Val Ala Pro Pro	
20 25 30	
cag aaa cac gaa gtt cgt atc aag att ctc ttc act tct ctc tgt cac	144
Gln Lys His Glu Val Arg Ile Lys Ile Leu Phe Thr Ser Leu Cys His	
35 40 45	
acc gat gtt tac ttc tgg gaa gct aag gga caa aca ccg ttg ttt cca	192
Thr Asp Val Tyr Phe Trp Glu Ala Lys Gly Gln Thr Pro Leu Phe Pro	
50 55 60	
cgt atc ttc ggc cat gaa gct gga ggg att gtt gag agt gtt gga gaa	240
Arg Ile Phe Gly His Glu Ala Gly Gly Ile Val Glu Ser Val Gly Glu	
65 70 75 80	
gga gtg act gat ctt cag cca gga gat cat gtg ttg ccg atc ttt acc	288
Gly Val Thr Asp Leu Gln Pro Gly Asp His Val Leu Pro Ile Phe Thr	
85 90 95	
gga gaa tgt gga gat tgt cgt cat tgc cag tcg gag gaa tca aac atg	336
Gly Glu Cys Gly Asp Cys Arg His Cys Gln Ser Glu Glu Ser Asn Met	
100 105 110	
tgt gat ctt ctc agg atc aac aca gag cga gga ggt atg att cac gat	384
Cys Asp Leu Leu Arg Ile Asn Thr Glu Arg Gly Gly Met Ile His Asp	
115 120 125	
ggt gaa tct aga ttc tcc att aat ggc aaa cca atc tac cat ttc ctt	432
Gly Glu Ser Arg Phe Ser Ile Asn Gly Lys Pro Ile Tyr His Phe Leu	
130 135 140	
ggg acg tcc acg ttc agt gag tac act gtg gtt cac tct ggt cag gtc	480
Gly Thr Ser Thr Phe Ser Glu Tyr Thr Val Val His Ser Gly Gln Val	
145 150 155 160	
gct aag atc aat ccg gat gct cct ctt gac aag gtc tgt att gtc agt	528
Ala Lys Ile Asn Pro Asp Ala Pro Leu Asp Lys Val Cys Ile Val Ser	
165 170 175	
tgt ggt ttg tct act ggg tta gga gca act ttg aat gtg gct aaa ccc	576
Cys Gly Leu Ser Thr Gly Leu Gly Ala Thr Leu Asn Val Ala Lys Pro	
180 185 190	
aag aaa ggt caa agt gtt gcc att ttt ggt ctt ggt gct gtt ggt tta	624
Lys Lys Gly Gln Ser Val Ala Ile Phe Gly Leu Gly Ala Val Gly Leu	
195 200 205	
ggc gct gca gaa ggt gct aga atc gct ggt gct tct agg atc atc ggt	672
Gly Ala Ala Glu Gly Ala Arg Ile Ala Gly Ala Ser Arg Ile Ile Gly	
210 215 220	
gtt gat ttt aac tct aaa aga ttc gac caa gct aag gaa ttc ggt gtg	720
Val Asp Phe Asn Ser Lys Arg Phe Asp Gln Ala Lys Glu Phe Gly Val	
225 230 235 240	
acc gag tgt gtg aac ccg aaa gac cat gac aag cca att caa cag gtg	768
Thr Glu Cys Val Asn Pro Lys Asp His Asp Lys Pro Ile Gln Gln Val	
245 250 255	
atc gct gag atg acg gat ggt ggg gtg gac agg agt gtg gaa tgc acc	816
Ile Ala Glu Met Thr Asp Gly Gly Val Asp Arg Ser Val Glu Cys Thr	
260 265 270	

gga agc gtt cag gcc atg att caa gca ttt gaa tgt gtc cac gat ggc 864  
 Gly Ser Val Gln Ala Met Ile Gln Ala Phe Glu Cys Val His Asp Gly  
 275 280 285  
 tgg ggt gtt gca gtg ctg gtg ggt gtg cca agc aaa gac gat gcc ttc 912  
 Trp Gly Val Ala Val Leu Val Gly Val Pro Ser Lys Asp Asp Ala Phe  
 290 295 300  
 aag act cat ccg atg aat ttc ttg aat gag agg act ctt aag ggt act 960  
 Lys Thr His Pro Met Asn Phe Leu Asn Glu Arg Thr Leu Lys Gly Thr  
 305 310 315 320  
 ttc ttc ggg aac tac aaa ccc aaa act gac att ccc ggg gtt gtg gaa 1008  
 Phe Phe Gly Asn Tyr Lys Pro Lys Thr Asp Ile Pro Gly Val Val Glu  
 325 330 335  
 aag tac atg aac aag gag ctg gag ctt gag aaa ttc atc act cac aca 1056  
 Lys Tyr Met Asn Lys Glu Leu Glu Leu Glu Lys Phe Ile Thr His Thr  
 340 345 350  
 gtg cca ttc tcg gaa atc aac aag gcc ttt gat tac atg ctg aag gga 1104  
 Val Pro Phe Ser Glu Ile Asn Lys Ala Phe Asp Tyr Met Leu Lys Gly  
 355 360 365  
 gag agt att cgt tgc atc atc acc atg ggt gct tga 1140  
 Glu Ser Ile Arg Cys Ile Ile Thr Met Gly Ala  
 370 375

&lt;210&gt; 42

&lt;211&gt; 379

&lt;212&gt; PRT

&lt;213&gt; Arabidopsis thaliana

&lt;400&gt; 42

Met Ser Thr Thr Gly Gln Ile Ile Arg Cys Lys Ala Ala Val Ala Trp  
 1 5 10 15  
 Glu Ala Gly Lys Pro Leu Val Ile Glu Glu Val Glu Val Ala Pro Pro  
 20 25 30  
 Gln Lys His Glu Val Arg Ile Lys Ile Leu Phe Thr Ser Leu Cys His  
 35 40 45  
 Thr Asp Val Tyr Phe Trp Glu Ala Lys Gly Gln Thr Pro Leu Phe Pro  
 50 55 60  
 Arg Ile Phe Gly His Glu Ala Gly Gly Ile Val Glu Ser Val Gly Glu  
 65 70 75 80  
 Gly Val Thr Asp Leu Gln Pro Gly Asp His Val Leu Pro Ile Phe Thr  
 85 90 95  
 Gly Glu Cys Gly Asp Cys Arg His Cys Gln Ser Glu Glu Ser Asn Met  
 100 105 110  
 Cys Asp Leu Leu Arg Ile Asn Thr Glu Arg Gly Gly Met Ile His Asp  
 115 120 125  
 Gly Glu Ser Arg Phe Ser Ile Asn Gly Lys Pro Ile Tyr His Phe Leu  
 130 135 140  
 Gly Thr Ser Thr Phe Ser Glu Tyr Thr Val Val His Ser Gly Gln Val  
 145 150 155 160  
 Ala Lys Ile Asn Pro Asp Ala Pro Leu Asp Lys Val Cys Ile Val Ser  
 165 170 175  
 Cys Gly Leu Ser Thr Gly Leu Gly Ala Thr Leu Asn Val Ala Lys Pro  
 180 185 190



Lys Lys Gly Gln Ser Val Ala Ile Phe Gly Leu Gly Ala Val Gly Leu  
 195 200 205  
 Gly Ala Ala Glu Gly Ala Arg Ile Ala Gly Ala Ser Arg Ile Ile Gly  
 210 215 220  
 Val Asp Phe Asn Ser Lys Arg Phe Asp Gln Ala Lys Glu Phe Gly Val  
 225 230 235 240  
 Thr Glu Cys Val Asn Pro Lys Asp His Asp Lys Pro Ile Gln Gln Val  
 245 250 255  
 Ile Ala Glu Met Thr Asp Gly Gly Val Asp Arg Ser Val Glu Cys Thr  
 260 265 270  
 Gly Ser Val Gln Ala Met Ile Gln Ala Phe Glu Cys Val His Asp Gly  
 275 280 285  
 Trp Gly Val Ala Val Leu Val Gly Val Pro Ser Lys Asp Asp Ala Phe  
 290 295 300  
 Lys Thr His Pro Met Asn Phe Leu Asn Glu Arg Thr Leu Lys Gly Thr  
 305 310 315 320  
 Phe Phe Gly Asn Tyr Lys Pro Lys Thr Asp Ile Pro Gly Val Val Glu  
 325 330 335  
 Lys Tyr Met Asn Lys Glu Leu Glu Leu Glu Lys Phe Ile Thr His Thr  
 340 345 350  
 Val Pro Phe Ser Glu Ile Asn Lys Ala Phe Asp Tyr Met Leu Lys Gly  
 355 360 365  
 Glu Ser Ile Arg Cys Ile Ile Thr Met Gly Ala  
 370 375

&lt;210&gt; 43

&lt;211&gt; 1140

&lt;212&gt; DNA

&lt;213&gt; Hordeum vulgare

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1137)

&lt;223&gt; coding for alcohol dehydrogenase

&lt;400&gt; 43

atg gcg acg gcc ggc aag gtg atc aag tgc aaa gcc gcg gtg gcg tgg 48

Met Ala Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Ala Trp

1

5

10

15

gag gcc ggg aag ccg ctg acc atg gag gag gtg gag gtg gcg ccg ccg 96

Glu Ala Gly Lys Pro Leu Thr Met Glu Glu Val Glu Val Ala Pro Pro

20

25

30

cag gcc atg gag gtg cgc gtc aag atc ctc ttc acc tcc ctc tgc cac 144

Gln Ala Met Glu Val Arg Val Lys Ile Leu Phe Thr Ser Leu Cys His

35

40

45

acc gac gtc tac ttc tgg gag gcc aag ggg cag acc ccc atg ttc cct 192

Thr Asp Val Tyr Phe Trp Glu Ala Lys Gly Gln Thr Pro Met Phe Pro

50

55

60

cgg atc ttc ggc cat gaa gct gga ggc ata gtg gag agt gtt gga gag 240

Arg Ile Phe Gly His Glu Ala Gly Gly Ile Val Glu Ser Val Gly Glu

65

70

75

80

ggc	gtg	act	gat	gtt	gcc	cct	ggt	gac	cac	gtc	ctc	cct	gtg	ttc	act	288
Gly	Val	Thr	Asp	Val	Ala	Pro	Gly	Asp	His	Val	Leu	Pro	Val	Phe	Thr	
				85					90					95		
ggg	gag	tgt	aag	gaa	tgc	cca	cat	tgc	aag	tct	gcg	gag	agc	aac	atg	336
Gly	Glu	Cys	Lys	Glu	Cys	Pro	His	Cys	Lys	Ser	Ala	Glu	Ser	Asn	Met	
			100					105					110			
tgt	gat	ctg	ctc	agg	atc	aac	acc	gac	aga	ggt	gtg	atg	atc	ggg	gat	384
Cys	Asp	Leu	Leu	Arg	Ile	Asn	Thr	Asp	Arg	Gly	Val	Met	Ile	Gly	Asp	
			115				120					125				
ggc	aag	tgc	cgc	ttc	tct	att	ggc	ggc	aag	ccg	att	tac	cat	ttc	gta	432
Gly	Lys	Ser	Arg	Phe	Ser	Ile	Gly	Gly	Lys	Pro	Ile	Tyr	His	Phe	Val	
	130					135					140					
ggg	act	tcc	acc	ttc	agt	gag	tac	act	gtc	atg	cat	gtc	ggt	tgt	gtt	480
Gly	Thr	Ser	Thr	Phe	Ser	Glu	Tyr	Thr	Val	Met	His	Val	Gly	Cys	Val	
	145				150					155					160	
gcc	aag	atc	aac	cct	gag	gct	ccc	ctt	gat	aaa	gtc	tgt	gtt	ctt	agc	528
Ala	Lys	Ile	Asn	Pro	Glu	Ala	Pro	Leu	Asp	Lys	Val	Cys	Val	Leu	Ser	
				165					170					175		
tgt	ggt	att	tgc	act	ggt	ctt	ggc	gcg	tca	att	aat	gtt	gca	aaa	cca	576
Cys	Gly	Ile	Cys	Thr	Gly	Leu	Gly	Ala	Ser	Ile	Asn	Val	Ala	Lys	Pro	
			180					185					190			
cca	aag	ggt	tcc	aca	gtg	gcg	ata	ttt	ggg	cta	gga	gct	gtt	ggc	ctt	624
Pro	Lys	Gly	Ser	Thr	Val	Ala	Ile	Phe	Gly	Leu	Gly	Ala	Val	Gly	Leu	
			195				200					205				
gct	gct	gca	gaa	ggt	gca	agg	att	gca	ggt	gca	tca	agg	atc	att	ggt	672
Ala	Ala	Ala	Glu	Gly	Ala	Arg	Ile	Ala	Gly	Ala	Ser	Arg	Ile	Ile	Gly	
	210					215					220					
gtt	gac	ctg	aac	gcc	agc	aga	ttt	gaa	gag	gct	agg	aag	ttt	ggc	tgc	720
Val	Asp	Leu	Asn	Ala	Ser	Arg	Phe	Glu	Glu	Ala	Arg	Lys	Phe	Gly	Cys	
	225				230				235					240		
acg	gaa	ttt	gtg	aac	ccg	aaa	gat	cac	acc	aag	cca	gtt	cag	cag	gtg	768
Thr	Glu	Phe	Val	Asn	Pro	Lys	Asp	His	Thr	Lys	Pro	Val	Gln	Gln	Val	
				245				250						255		
ctc	gct	gac	atg	aca	aat	ggc	gga	gtt	gac	cgc	agt	gtt	gag	tgc	act	816
Leu	Ala	Asp	Met	Thr	Asn	Gly	Gly	Val	Asp	Arg	Ser	Val	Glu	Cys	Thr	
			260					265					270			
ggc	aac	gtc	aat	gct	atg	ata	caa	gca	ttt	gaa	tgt	gtt	cat	gat	ggc	864
Gly	Asn	Val	Asn	Ala	Met	Ile	Gln	Ala	Phe	Glu	Cys	Val	His	Asp	Gly	
	275						280					285				
tgg	ggt	gta	gct	gtg	ctg	gtg	ggt	gtg	cca	cac	aag	gac	gct	gaa	ttc	912
Trp	Gly	Val	Ala	Val	Leu	Val	Gly	Val	Pro	His	Lys	Asp	Ala	Glu	Phe	
	290					295					300					
aag	acc	cac	ccg	atg	aac	ttc	ctg	aat	gag	agg	acc	ctg	aag	ggc	acc	960
Lys	Thr	His	Pro	Met	Asn	Phe	Leu	Asn	Glu	Arg	Thr	Leu	Lys	Gly	Thr	
	305				310				315					320		
ttc	ttc	ggt	aac	ttc	aag	ccg	cgc	act	gac	ctg	ccc	aat	gtc	gtg	gag	1008
Phe	Phe	Gly	Asn	Phe	Lys	Pro	Arg	Thr	Asp	Leu	Pro	Asn	Val	Val	Glu	
			325					330						335		
atg	tac	atg	aag	aag	gag	ctg	gag	gtg	gag	aag	ttc	atc	aca	cac	agc	1056
Met	Tyr	Met	Lys	Lys	Glu	Leu	Glu	Val	Glu	Lys	Phe	Ile	Thr	His	Ser	
			340					345					350			

gtg ccg ttc tcg gag ata aac aag gcc ttc gac ctt atg gcg aag ggg 1104  
Val Pro Phe Ser Glu Ile Asn Lys Ala Phe Asp Leu Met Ala Lys Gly  
355 360 365

gag ggc atc cgt tgc atc atc cgc atg gac aac tag 1140  
Glu Gly Ile Arg Cys Ile Ile Arg Met Asp Asn  
370 375

<210> 44

<211> 379

<212> PRT

<213> Hordeum vulgare

<400> 44

Met 1	Ala	Thr	Ala	Gly 5	Lys	Val	Ile	Lys	Cys 10	Lys	Ala	Ala	Val	Ala 15	Trp
Glu	Ala	Gly	Lys 20	Pro	Leu	Thr	Met	Glu 25	Glu	Val	Glu	Val	Ala 30	Pro	Pro
Gln	Ala	Met 35	Glu	Val	Arg	Val	Lys 40	Ile	Leu	Phe	Thr	Ser 45	Leu	Cys	His
Thr	Asp 50	Val	Tyr	Phe	Trp	Glu 55	Ala	Lys	Gly	Gln	Thr 60	Pro	Met	Phe	Pro
Arg 65	Ile	Phe	Gly	His 70	Glu	Ala	Gly	Gly	Ile	Val 75	Glu	Ser	Val	Gly	Glu 80
Gly	Val	Thr	Asp	Val 85	Ala	Pro	Gly	Asp	His 90	Val	Leu	Pro	Val	Phe 95	Thr
Gly	Glu	Cys	Lys 100	Glu	Cys	Pro	His	Cys 105	Lys	Ser	Ala	Glu	Ser	Asn	Met
Cys	Asp 115	Leu	Leu	Arg	Ile	Asn	Thr 120	Asp	Arg	Gly	Val	Met 125	Ile	Gly	Asp
Gly 130	Lys	Ser	Arg	Phe	Ser 135	Ile	Gly	Gly	Lys	Pro	Ile 140	Tyr	His	Phe	Val
Gly 145	Thr	Ser	Thr	Phe 150	Ser	Glu	Tyr	Thr	Val	Met 155	His	Val	Gly	Cys	Val 160
Ala	Lys	Ile	Asn 165	Pro	Glu	Ala	Pro	Leu	Asp 170	Lys	Val	Cys	Val	Leu 175	Ser
Cys	Gly	Ile	Cys 180	Thr	Gly	Leu	Gly	Ala 185	Ser	Ile	Asn	Val	Ala 190	Lys	Pro
Pro	Lys 195	Gly	Ser	Thr	Val	Ala	Ile 200	Phe	Gly	Leu	Gly	Ala 205	Val	Gly	Leu
Ala 210	Ala	Ala	Glu	Gly	Ala 215	Arg	Ile	Ala	Gly	Ala	Ser 220	Arg	Ile	Ile	Gly
Val 225	Asp	Leu	Asn	Ala 230	Ser	Arg	Phe	Glu	Glu	Ala 235	Arg	Lys	Phe	Gly	Cys 240
Thr	Glu	Phe	Val	Asn 245	Pro	Lys	Asp	His	Thr 250	Lys	Pro	Val	Gln	Gln 255	Val
Leu	Ala	Asp	Met 260	Thr	Asn	Gly	Gly	Val	Asp 265	Arg	Ser	Val	Glu	Cys	Thr
Gly	Asn 275	Val	Asn	Ala	Met	Ile	Gln 280	Ala	Phe	Glu	Cys 285	Val	His	Asp	Gly
Trp 290	Gly	Val	Ala	Val	Leu 295	Val	Gly	Val	Pro	His	Lys 300	Asp	Ala	Glu	Phe

Lys Thr His Pro Met Asn Phe Leu Asn Glu Arg Thr Leu Lys Gly Thr  
 305 310 315 320  
 Phe Phe Gly Asn Phe Lys Pro Arg Thr Asp Leu Pro Asn Val Val Glu  
 325 330 335  
 Met Tyr Met Lys Lys Glu Leu Glu Val Glu Lys Phe Ile Thr His Ser  
 340 345 350  
 Val Pro Phe Ser Glu Ile Asn Lys Ala Phe Asp Leu Met Ala Lys Gly  
 355 360 365  
 Glu Gly Ile Arg Cys Ile Ile Arg Met Asp Asn  
 370 375

&lt;210&gt; 45

&lt;211&gt; 1140

&lt;212&gt; DNA

&lt;213&gt; Oryza sativa

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1137)

&lt;223&gt; coding for alcohol dehydrogenase

&lt;400&gt; 45

atg gcg acc gca ggg aag gtg atc aag tgc aaa gcg gcg gtg gca tgg 48  
 Met Ala Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Ala Trp  
 1 5 10 15  
 gag gcc gcg aag ccg ctg gtg atc gag gag gtg gag gtg gcg ccg ccg 96  
 Glu Ala Ala Lys Pro Leu Val Ile Glu Glu Val Glu Val Ala Pro Pro  
 20 25 30  
 cag gcc atg gag gtg cgc gtc aag atc ctc ttc acc tcg ctc tgc cac 144  
 Gln Ala Met Glu Val Arg Val Lys Ile Leu Phe Thr Ser Leu Cys His  
 35 40 45  
 acc gac gtc tac ttc tgg gag gcc aag gga cag act ccc gtg ttc cct 192  
 Thr Asp Val Tyr Phe Trp Glu Ala Lys Gly Gln Thr Pro Val Phe Pro  
 50 55 60  
 cgg atc ttc ggc cat gaa gct gga ggt att gtg gag agt gtt gga gag 240  
 Arg Ile Phe Gly His Glu Ala Gly Gly Ile Val Glu Ser Val Gly Glu  
 65 70 75 80  
 ggt gtg act gat ctt gcc cct ggt gac cat gtt ctc cct gtg ttc act 288  
 Gly Val Thr Asp Leu Ala Pro Gly Asp His Val Leu Pro Val Phe Thr  
 85 90 95  
 ggg gag tgc aag gag tgt gcc cac tgc aag tca gca gag agc aac atg 336  
 Gly Glu Cys Lys Glu Cys Ala His Cys Lys Ser Ala Glu Ser Asn Met  
 100 105 110  
 tgt gat ctg ctc agg atc aac act gac agg ggt gtg atg att ggt gat 384  
 Cys Asp Leu Leu Arg Ile Asn Thr Asp Arg Gly Val Met Ile Gly Asp  
 115 120 125  
 ggc aaa tca cgc ttt tcc atc aac ggg aag ccc att tac cat ttc gtc 432  
 Gly Lys Ser Arg Phe Ser Ile Asn Gly Lys Pro Ile Tyr His Phe Val  
 130 135 140  
 ggg act tcg acc ttc agc gag tac act gtc atg cat gtt ggt tgc gtt 480  
 Gly Thr Ser Thr Phe Ser Glu Tyr Thr Val Met His Val Gly Cys Val  
 145 150 155 160

gcg aag atc aac ccg gca gct cca ctt gat aaa gtt tgc gtt ctt agc 528  
 Ala Lys Ile Asn Pro Ala Ala Pro Leu Asp Lys Val Cys Val Leu Ser  
 165 170 175

tgt ggt att tct act ggt ctt ggt gct aca atc aat gtg gca aag cca 576  
 Cys Gly Ile Ser Thr Gly Leu Gly Ala Thr Ile Asn Val Ala Lys Pro  
 180 185 190

cca aag ggt tgc acg gtg gcg ata ttt ggt cta gga gct gta ggc ctt 624  
 Pro Lys Gly Ser Thr Val Ala Ile Phe Gly Leu Gly Ala Val Gly Leu  
 195 200 205

gct gcc gca gaa ggt gca agg att gca gga gcg tca agg atc att ggc 672  
 Ala Ala Ala Glu Gly Ala Arg Ile Ala Gly Ala Ser Arg Ile Ile Gly  
 210 215 220

att gac ctg aac gcc aac aga ttt gaa gaa gct agg aaa ttt ggt tgc 720  
 Ile Asp Leu Asn Ala Asn Arg Phe Glu Glu Ala Arg Lys Phe Gly Cys  
 225 230 235 240

act gaa ttt gtg aac cca aag gac cat gac aag cca gtt cag cag gta 768  
 Thr Glu Phe Val Asn Pro Lys Asp His Asp Lys Pro Val Gln Gln Val  
 245 250 255

ctt gct gag atg acc aat ggc gga gtt gac cgc agc gtt gaa tgc act 816  
 Leu Ala Glu Met Thr Asn Gly Gly Val Asp Arg Ser Val Glu Cys Thr  
 260 265 270

ggc aac atc aac gcc atg atc caa gca ttt gaa tgt gtt cat gat ggc 864  
 Gly Asn Ile Asn Ala Met Ile Gln Ala Phe Glu Cys Val His Asp Gly  
 275 280 285

tgg ggt gtt gct gtt ttg gtc ggc gtg cca cac aag gac gcc gag ttc 912  
 Trp Gly Val Ala Val Leu Val Gly Val Pro His Lys Asp Ala Glu Phe  
 290 295 300

aag acc cac ccg atg aac ttc ctg aac gag agg act ctc aag gga acc 960  
 Lys Thr His Pro Met Asn Phe Leu Asn Glu Arg Thr Leu Lys Gly Thr  
 305 310 315 320

ttc ttc ggc aac tac aag cca cgc acc gat ctg ccc aac gtc gtc gag 1008  
 Phe Phe Gly Asn Tyr Lys Pro Arg Thr Asp Leu Pro Asn Val Val Glu  
 325 330 335

ctc tac atg aag aag gag ctg gag gtg gag aag ttc atc aca cac agc 1056  
 Leu Tyr Met Lys Lys Glu Leu Glu Val Glu Lys Phe Ile Thr His Ser  
 340 345 350

gtg ccg ttc tgc gag atc aac acg gcg ttc gac ctg atg cac aag ggc 1104  
 Val Pro Phe Ser Glu Ile Asn Thr Ala Phe Asp Leu Met His Lys Gly  
 355 360 365

gag ggc atc cgc tgc atc atc cgc atg gag aac tga 1140  
 Glu Gly Ile Arg Cys Ile Ile Arg Met Glu Asn  
 370 375

<210> 46  
 <211> 379  
 <212> PRT  
 <213> Oryza sativa  
 <400> 46  
 Met Ala Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Ala Trp  
 1 5 10 15  
 Glu Ala Ala Lys Pro Leu Val Ile Glu Glu Val Glu Val Ala Pro Pro  
 20 25 30

Gln Ala Met Glu Val Arg Val Lys Ile Leu Phe Thr Ser Leu Cys His  
                   35                                  40                                  45  
 Thr Asp Val Tyr Phe Trp Glu Ala Lys Gly Gln Thr Pro Val Phe Pro  
                   50                                  55                                  60  
 Arg Ile Phe Gly His Glu Ala Gly Gly Ile Val Glu Ser Val Gly Glu  
                   65                                  70                                  75                                  80  
 Gly Val Thr Asp Leu Ala Pro Gly Asp His Val Leu Pro Val Phe Thr  
                                   85                                  90                                  95  
 Gly Glu Cys Lys Glu Cys Ala His Cys Lys Ser Ala Glu Ser Asn Met  
                                   100                                  105                                  110  
 Cys Asp Leu Leu Arg Ile Asn Thr Asp Arg Gly Val Met Ile Gly Asp  
                   115                                  120                                  125  
 Gly Lys Ser Arg Phe Ser Ile Asn Gly Lys Pro Ile Tyr His Phe Val  
                   130                                  135                                  140  
 Gly Thr Ser Thr Phe Ser Glu Tyr Thr Val Met His Val Gly Cys Val  
                   145                                  150                                  155                                  160  
 Ala Lys Ile Asn Pro Ala Ala Pro Leu Asp Lys Val Cys Val Leu Ser  
                                   165                                  170                                  175  
 Cys Gly Ile Ser Thr Gly Leu Gly Ala Thr Ile Asn Val Ala Lys Pro  
                                   180                                  185                                  190  
 Pro Lys Gly Ser Thr Val Ala Ile Phe Gly Leu Gly Ala Val Gly Leu  
                   195                                  200                                  205  
 Ala Ala Ala Glu Gly Ala Arg Ile Ala Gly Ala Ser Arg Ile Ile Gly  
                   210                                  215                                  220  
 Ile Asp Leu Asn Ala Asn Arg Phe Glu Glu Ala Arg Lys Phe Gly Cys  
                   225                                  230                                  235                                  240  
 Thr Glu Phe Val Asn Pro Lys Asp His Asp Lys Pro Val Gln Gln Val  
                                   245                                  250                                  255  
 Leu Ala Glu Met Thr Asn Gly Gly Val Asp Arg Ser Val Glu Cys Thr  
                   260                                  265                                  270  
 Gly Asn Ile Asn Ala Met Ile Gln Ala Phe Glu Cys Val His Asp Gly  
                   275                                  280                                  285  
 Trp Gly Val Ala Val Leu Val Gly Val Pro His Lys Asp Ala Glu Phe  
                   290                                  295                                  300  
 Lys Thr His Pro Met Asn Phe Leu Asn Glu Arg Thr Leu Lys Gly Thr  
                   305                                  310                                  315                                  320  
 Phe Phe Gly Asn Tyr Lys Pro Arg Thr Asp Leu Pro Asn Val Val Glu  
                                   325                                  330                                  335  
 Leu Tyr Met Lys Lys Glu Leu Glu Val Glu Lys Phe Ile Thr His Ser  
                   340                                  345                                  350  
 Val Pro Phe Ser Glu Ile Asn Thr Ala Phe Asp Leu Met His Lys Gly  
                   355                                  360                                  365  
 Glu Gly Ile Arg Cys Ile Ile Arg Met Glu Asn  
                   370                                  375

&lt;210&gt; 47

&lt;211&gt; 1140

&lt;212&gt; DNA

&lt;213&gt; Zea mays

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(1137)

&lt;223&gt; coding for alcohol dehydrogenase

&lt;400&gt; 47

atg gcg acc gcg ggg aag gtg atc aag tgc aaa gct gcg gtg gca tgg	48
Met Ala Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Ala Trp	
1 5 10 15	
gag gcc ggc aag cca ctg tcg atc gag gag gtg gag gta gcg cct ccg	96
Glu Ala Gly Lys Pro Leu Ser Ile Glu Glu Val Glu Val Ala Pro Pro	
20 25 30	
cag gcc atg gag gtg cgc gtc aag atc ctc ttc acc tcg ctc tgc cac	144
Gln Ala Met Glu Val Arg Val Lys Ile Leu Phe Thr Ser Leu Cys His	
35 40 45	
acc gac gtc tac ttc tgg gag gcc aag ggg cag act ccc gtg ttc cct	192
Thr Asp Val Tyr Phe Trp Glu Ala Lys Gly Gln Thr Pro Val Phe Pro	
50 55 60	
cgg atc ttt ggc cat gag gct gga ggt atc ata gag agt gtt gga gag	240
Arg Ile Phe Gly His Glu Ala Gly Gly Ile Ile Glu Ser Val Gly Glu	
65 70 75 80	
ggt gtg act gac gta gct ccg ggc gac cat gtc ctt cct gtg ttc act	288
Gly Val Thr Asp Val Ala Pro Gly Asp His Val Leu Pro Val Phe Thr	
85 90 95	
ggg gag tgc aag gag tgc gcc cac tgc aag tcg gca gag agc aac atg	336
Gly Glu Cys Lys Glu Cys Ala His Cys Lys Ser Ala Glu Ser Asn Met	
100 105 110	
tgt gat ttg ctc agg atc aac act gac cgc ggt gtg atg att ggc gat	384
Cys Asp Leu Leu Arg Ile Asn Thr Asp Arg Gly Val Met Ile Gly Asp	
115 120 125	
ggc aag tcg cgg ttt tca atc aat ggg aag cct atc tac cac ttt gtt	432
Gly Lys Ser Arg Phe Ser Ile Asn Gly Lys Pro Ile Tyr His Phe Val	
130 135 140	
ggg act tcc acc ttc agc gag tac acc gtc atg cat gtc ggt tgt gtt	480
Gly Thr Ser Thr Phe Ser Glu Tyr Thr Val Met His Val Gly Cys Val	
145 150 155 160	
gca aag atc aac cct cag gct ccc ctt gat aaa gtt tgc gtc ctt agc	528
Ala Lys Ile Asn Pro Gln Ala Pro Leu Asp Lys Val Cys Val Leu Ser	
165 170 175	
tgt ggt att tct act ggt ctt ggt gca tca att aat gtt gca aaa cct	576
Cys Gly Ile Ser Thr Gly Leu Gly Ala Ser Ile Asn Val Ala Lys Pro	
180 185 190	
ccg aag ggt tcg aca gtg gct gtt ttc ggt tta gga gcc gtt ggt ctt	624
Pro Lys Gly Ser Thr Val Ala Val Phe Gly Leu Gly Ala Val Gly Leu	
195 200 205	
gcc gct gca gaa ggt gca agg att gct gga gcg tca agg atc att ggt	672
Ala Ala Ala Glu Gly Ala Arg Ile Ala Gly Ala Ser Arg Ile Ile Gly	
210 215 220	
gtc gac ctg aac ccc agc aga ttc gaa gaa gct agg aag ttc ggt tgc	720
Val Asp Leu Asn Pro Ser Arg Phe Glu Glu Ala Arg Lys Phe Gly Cys	
225 230 235 240	

act gaa ttt gtg aac cca aaa gac cac aac aag ccg gtg cag gag gta 768  
 Thr Glu Phe Val Asn Pro Lys Asp His Asn Lys Pro Val Gln Glu Val  
 245 250 255  
 ctt gct gag atg acc aac gga ggg gtc gac cgc agc gtg gaa tgc act 816  
 Leu Ala Glu Met Thr Asn Gly Gly Val Asp Arg Ser Val Glu Cys Thr  
 260 265 270  
 ggc aac atc aat gct atg atc caa gct ttc gaa tgt gtt cat gat ggc 864  
 Gly Asn Ile Asn Ala Met Ile Gln Ala Phe Glu Cys Val His Asp Gly  
 275 280 285  
 tgg ggt gtt gcc gtg ctg gtg ggt gtg ccg cat aag gac gct gag ttc 912  
 Trp Gly Val Ala Val Leu Val Gly Val Pro His Lys Asp Ala Glu Phe  
 290 295 300  
 aag acc cac ccg atg aac ttc ctg aac gaa agg acc ctg aag ggg acc 960  
 Lys Thr His Pro Met Asn Phe Leu Asn Glu Arg Thr Leu Lys Gly Thr  
 305 310 315 320  
 ttc ttt ggc aac tat aag cca cgc act gat ctg cca aat gtg gtg gag 1008  
 Phe Phe Gly Asn Tyr Lys Pro Arg Thr Asp Leu Pro Asn Val Val Glu  
 325 330 335  
 ctg tac atg aaa aag gag ctg gag gtg gag aag ttc atc acg cac agc 1056  
 Leu Tyr Met Lys Lys Glu Leu Glu Val Glu Lys Phe Ile Thr His Ser  
 340 345 350  
 gtc ccg ttc gcg gag atc aac aag gcg ttc aac ctg atg gcc aag ggg 1104  
 Val Pro Phe Ala Glu Ile Asn Lys Ala Phe Asn Leu Met Ala Lys Gly  
 355 360 365  
 gag ggc atc cgc tgc atc atc cgc atg gag aac tag 1140  
 Glu Gly Ile Arg Cys Ile Ile Arg Met Glu Asn  
 370 375  
 <210> 48  
 <211> 379  
 <212> PRT  
 <213> Zea mays  
 <400> 48  
 Met Ala Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Ala Trp  
 1 5 10 15  
 Glu Ala Gly Lys Pro Leu Ser Ile Glu Glu Val Glu Val Ala Pro Pro  
 20 25 30  
 Gln Ala Met Glu Val Arg Val Lys Ile Leu Phe Thr Ser Leu Cys His  
 35 40 45  
 Thr Asp Val Tyr Phe Trp Glu Ala Lys Gly Gln Thr Pro Val Phe Pro  
 50 55 60  
 Arg Ile Phe Gly His Glu Ala Gly Gly Ile Ile Glu Ser Val Gly Glu  
 65 70 75 80  
 Gly Val Thr Asp Val Ala Pro Gly Asp His Val Leu Pro Val Phe Thr  
 85 90 95  
 Gly Glu Cys Lys Glu Cys Ala His Cys Lys Ser Ala Glu Ser Asn Met  
 100 105 110  
 Cys Asp Leu Leu Arg Ile Asn Thr Asp Arg Gly Val Met Ile Gly Asp  
 115 120 125  
 Gly Lys Ser Arg Phe Ser Ile Asn Gly Lys Pro Ile Tyr His Phe Val  
 130 135 140



Gly	Thr	Ser	Thr	Phe	Ser	Glu	Tyr	Thr	Val	Met	His	Val	Gly	Cys	Val	
145					150					155					160	
Ala	Lys	Ile	Asn	Pro	Gln	Ala	Pro	Leu	Asp	Lys	Val	Cys	Val	Leu	Ser	
			165					170						175		
Cys	Gly	Ile	Ser	Thr	Gly	Leu	Gly	Ala	Ser	Ile	Asn	Val	Ala	Lys	Pro	
			180					185					190			
Pro	Lys	Gly	Ser	Thr	Val	Ala	Val	Phe	Gly	Leu	Gly	Ala	Val	Gly	Leu	
		195					200					205				
Ala	Ala	Ala	Glu	Gly	Ala	Arg	Ile	Ala	Gly	Ala	Ser	Arg	Ile	Ile	Gly	
	210					215					220					
Val	Asp	Leu	Asn	Pro	Ser	Arg	Phe	Glu	Glu	Ala	Arg	Lys	Phe	Gly	Cys	
225					230					235					240	
Thr	Glu	Phe	Val	Asn	Pro	Lys	Asp	His	Asn	Lys	Pro	Val	Gln	Glu	Val	
				245					250					255		
Leu	Ala	Glu	Met	Thr	Asn	Gly	Gly	Val	Asp	Arg	Ser	Val	Glu	Cys	Thr	
			260					265					270			
Gly	Asn	Ile	Asn	Ala	Met	Ile	Gln	Ala	Phe	Glu	Cys	Val	His	Asp	Gly	
	275						280					285				
Trp	Gly	Val	Ala	Val	Leu	Val	Gly	Val	Pro	His	Lys	Asp	Ala	Glu	Phe	
	290					295					300					
Lys	Thr	His	Pro	Met	Asn	Phe	Leu	Asn	Glu	Arg	Thr	Leu	Lys	Gly	Thr	
305					310					315					320	
Phe	Phe	Gly	Asn	Tyr	Lys	Pro	Arg	Thr	Asp	Leu	Pro	Asn	Val	Val	Glu	
			325					330						335		
Leu	Tyr	Met	Lys	Lys	Glu	Leu	Glu	Val	Glu	Lys	Phe	Ile	Thr	His	Ser	
		340					345						350			
Val	Pro	Phe	Ala	Glu	Ile	Asn	Lys	Ala	Phe	Asn	Leu	Met	Ala	Lys	Gly	
	355						360					365				
Glu	Gly	Ile	Arg	Cys	Ile	Ile	Arg	Met	Glu	Asn						
	370					375										

&lt;210&gt; 49

&lt;211&gt; 505

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz: coding for  
sense RNA-fragment of E.coli codA gene

&lt;400&gt; 49

```

aagcttggct aacagtgtcg aataacgctt taaaacaat tattaacgcc cggttaccag 60
gcgaagaggg gctgtggcag attcatctgc aggacggaaa aatcagcgcc attgatgcgc 120
aatccggcgt gatgccata actgaaaaca gcctggatgc cgaacaagg ttagttatac 180
cgccgtttgt ggagccacat attcacctgg acaccacgca aaccgccgga caaccgaact 240
ggaatcagtc cggcacgctg tttgaaggca ttgaacgctg ggccgagcgc aaagcgttat 300
taaccatga cgatgtgaaa caacgcgcgc ggcaaagcgt gaaatggcag attgccaacg 360
gcattcagca tgtgcgtacc catgtcgatg tttcgatgc aacgctaact gcgctgaaag 420
caatgctgga agtgaagcag gaagtcgcgc cgtggattga tctgcaaata gtcgccttcc 480
ctcaggaagg gattttgtcg tcgac

```

&lt;210&gt; 50

&lt;211&gt; 27

<212> DNA  
 <213> Künstliche Sequenz  
 <220>  
 <223> Beschreibung der künstlichen Sequenz:  
 oligonucleotide primer  
 <400> 50  
 cgtgaatacgc gcgtggagtc g 21  
 <210> 51  
 <211> 26  
 <212> DNA  
 <213> Künstliche Sequenz  
 <220>  
 <223> Beschreibung der künstlichen Sequenz:  
 oligonucleotide primer  
 <400> 51  
 cggcaggata atcaggttgg 20  
 <210> 52  
 <211> 505  
 <212> DNA  
 <213> Künstliche Sequenz  
 <220>  
 <223> Beschreibung der künstlichen Sequenz: coding for  
 antisense RNA-fragment of E.coli codA gene  
 <400> 52  
 gaattcggct aacagtgtcg aataacgctt tacaacaat tattaacgcc cggttaccag 60  
 gcgaagaggg gctgtggcag attcatctgc aggacggaaa aatcagcgcc attgatgcgc 120  
 aatccggcgt gatgccata actgaaaaca gcctggatgc cgaacaaggc ttagttatac 180  
 cgccgtttgt ggagccacat attcacctgg acaccacgca aaccgcccga caaccgaact 240  
 ggaatcagtc cggcacgctg tttgaaggca ttgaacgctg ggccgagcgc aaagcggttat 300  
 taacccatga cgatgtgaaa caacgcgcgc ggcaaacgct gaaatggcag attgccaacg 360  
 gcattcagca tgtgcgtacc catgtcgatg ttccggatgc aacgctaact gcgctgaaag 420  
 caatgctgga agtgaagcag gaagtcgcgc cgtggattga tctgcaaata gtcgccttcc 480  
 ctcaggaagg gattttgtcg gatcc 505  
 <210> 53  
 <211> 27  
 <212> DNA  
 <213> Künstliche Sequenz  
 <220>  
 <223> Beschreibung der künstlichen Sequenz:  
 oligonucleotide primer  
 <400> 53  
 gtcaacgtaa ccaaccctgc 20  
 <210> 54  
 <211> 26  
 <212> DNA  
 <213> Künstliche Sequenz  
 <220>  
 <223> Beschreibung der künstlichen Sequenz:  
 oligonucleotide primer  
 <400> 54  
 ggatccgaca aaatcccttc ctgagg 26

&lt;210&gt; 55

&lt;211&gt; 5674

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz: vector  
construct pBluKS-nitP-STLS1-35S-T

&lt;400&gt; 55

```

ccagcttttg ttcccttttag tgaggggtaa tttcgagctt ggcgtaatca tgggtcatagc 60
tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca 120
taaagtgtaa agcctggggg gcctaagtga tgagctaact cacattaatt gcgttgcgct 180
cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac 240
gcgcgggggag aggcgggttg cgtattgggc gctcttcgcg ttcctcgctc actgactcgc 300
tgcgctcggg cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg 360
tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg 420
ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg 480
agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat 540
accaggcggt tccccctgga agctccctcg tgcgctctcc tgttcgacc ctgccgcta 600
ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct 660
gtaggtatct cagttcgggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc 720
ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aaccgggtaa 780
gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggatat 840
taggcgggtg tacagagttc ttgaagtggg ggcctaacta cggctacact agaaggacag 900
tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt 960
gatccggcaa acaaacacc cgtggtagcg gtgggttttt tgtttgcaag cagcagatta 1020
cgcgagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc 1080
agtggaaacg aaactcacgt taagggatct tgggtcatgag attatcaaaa aggatcttca 1140
cctagatcct tttaaattaa aaatgaagtt ttaaataaat ctaaagtata tatgagtaaa 1200
cttgggtctg cagttaccaaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat 1260
ttcggttcac catagttgcc tgactccccg tctgttagat aactacgata cgggaggggt 1320
taccatctgg cccagtgctg gcaatgatac cgcgagacc acgctcaccg gctccagatt 1380
tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtgggtcct gcaactttat 1440
ccgcctccat ccagtctatt aattgttgcc ggggaagctag agtaagtagt tcgccagtta 1500
atagtttgcg caacgttggt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg 1560
gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tccccatgt 1620
tgtgcaaaaa agcgggttagc tccttcgggtc ctccgatcgt tgtcagaagt aagttggccg 1680
cagtggtatc actcatgggt atggcagcac tgcataatc tcttactgtc atgccatccg 1740
taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc 1800
ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa 1860
ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac 1920
cgctgttgag atccagttcg atgtaacca ctcgtgcacc caactgatct tcagcatctt 1980
ttactttcac cagcgtttct gggtagcaaa aaacaggaag gcaaaatgcc gcaaaaaagg 2040
gaataaggcg gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa 2100
gcatttatca ggggtattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata 2160
aacaatatag ggttccgcgc acatttcccc gaaaagtgcc acctgacgcg ccctgtagcg 2220
gcgcattaag cgcggcgggt gtggtgggta cgcgcagcgt gaccgctaca cttgccagcg 2280
ccctagcgcc cgctcctttc gctttcttcc cttcctttct cgccacgttc gccggctttc 2340
cccgtaagc tctaaatcgg gggctccctt taggggtccg atttagtgct ttacggcacc 2400
tcgaccccaa aaaacttgat tagggtagtg gttcacgtag tgggccatcg ccctgataga 2460
cggtttttcg ccctttgacg ttggagtcca cgttctttta tagtggactc ttgttccaaa 2520
ctggaacaac actcaaccct atctcgggtc attcttttga tttataaggg attttgccga 2580
tttcggccta ttgggttaaaa aatgagctga ttttaacaaa atttaacgcg aattttaaca 2640
aaatattaac gcttacaatt tccattcgcc attcaggctg cgcaactgtt ggaaggcg 2700
atcgggtcgg gcctcttcgc tattacgcca cgtggcgaaa ggggatgtg ctgcaaggcg 2760
attaagttgg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga cggccagtga 2820
attgtaatac gactcactat agggcgaatt ggagctcgtc gagaccagat gttttacact 2880

```

```

tgaccgtaaa tgagcacccg aagaaaccgg tcacattcat ttcgaagggt gagaaagcgg 2940
aagatgactc aaacaagtaa tcggttgtga ttcgtcagtt catgtcactc ctatgaagga 3000
gtcaagttca aaatgttatg ttgagtttca aacttttatg ctaaactttt tttctttatt 3060
ttcgtaata atggaagaga accaattctc ttgtatctaa agattatcca tctatcatcc 3120
aatttgagtg ttcaattctg gatgttgtgt taccctacat tctacaacca tgtagccaat 3180
tattatgaat ctggccttga tttcagttgt gttcttttct tttttttctt tgcataattg 3240
catttagaat gtttaataat taagttactg tatttccaca tacattagtt ccaagaatat 3300
acatatatta atttattttt cttaaaaatg ttttggaatg actaatattg acaacgaaaa 3360
tagaagctat gctaaaccat tacgtatatg tgacttcaca tgttggtgtt ttacattccc 3420
tatatatatg gatggctgtc acaatcagaa acgtgatcga aaaaagacaa acagtgtttg 3480
cataaaaaga ctatttcgtt tcattgacaa tttgtgttta tttgtaaaga aaagtggcaa 3540
agtggaaatt gagttcctgc aagtaagaaa gatgaaataa aagacttgag tgtgtgtttt 3600
tttcttttat ctgaaagctg caatgaaata ttctaccaa gcccgtttga ttattaattg 3660
gggtttgggt tttctgatgc gaactaattg gttatataag aaactataca atccatgtta 3720
attcaaaaat tttgatttct cttgtaggaa tatgatttac tatatgagac tttcttttctg 3780
ccaataatag taaatccaaa gatatttgac cggaccacaaa cacattgatc tatttttttag 3840
tttattttaat ccagtttctc tgagataatt cattaaggaa aacttagtat taacccatcc 3900
taagattaaa taggagccaa actcacattt caaatattaa ataacataaa atggatttaa 3960
aaaatctata cgtcaaattt tatttatgac atttcttatt taaatttata tttaatgaaa 4020
tacagctaag acaaaccaaa aaaaaaatac tttctaagtg gtccaaaaca tcaattccgt 4080
tcaatattat taggtagaat cgtacgacca aaaaaggta ggtaatacgc aattagaac 4140
atatctataa catagtatat attattacct atttgagga atcaaaatgc atcaaatatg 4200
gatttaagga atccataaaa gaataaattc tacgggaaaa aaaatggaat aaattctttt 4260
aagtttttta tttgtttttt atttggtagt tctccatttt gttttatttc gtttggtttt 4320
attgtgtcca aatactttgt aaaccaccgt tgtaattctt aaacggggtt ttcacttctt 4380
ttttatattc agacataaag catcggtcgg tttaatcaat caatagattt tatttttctt 4440
ctcaattatt agtaggtttg atgtgaactt tacaaaaaaa acaaaaacaa atcaatgcag 4500
agaaaagaaa ccacgtgggc tagtcccacc ttgtttcatt tccaccacag gttcgatctt 4560
cgttaccgtc tccaatagga aaataaacgt gaccacaaaa aaaaaacaaa aaaaagtcta 4620
tatattgctt ctctcaagtc tctgagtgtc atgaacccaa gtaaaaaaca aagactcgac 4680
ctgcaggcat gcaagcttat cgtcgactac gtaagtttct gcttctacct ttgatata 4740
tataataatt atcattaatt agtagtaata taatatttca aatatttttt tcaaaataaa 4800
agaatgtagt atatagcaat tgcttttctg tagtttataa gtgtgtatat ttaattttat 4860
aacttttcta atatatgacc aaaatttgtt gatgtgcagg tatcaccgga tccatcgaat 4920
tcggtacgct gaaatcacca gtctctctct acaaatctat ctctctctat tttctccata 4980
aataatgtgt gagtagtttc ccgataaggg gaanttaggg ttcttatagg gtttcgctca 5040
tgtgttgagc atataagaaa cccttagtat gtatttgtat ttgtaaaata cttctatcaa 5100
taaaatttct aattcctaaa accaaaatcc agtactaaaa tccagatctc ctaaagtccc 5160
tatagatctt tgctcgtgaat ataaaccaga cacgagacga ctaaacctgg agcccagacg 5220
ccgttcgaag ctagaagtac cgcttaggca ggaggccgtt agggaaaaga tgctaaggca 5280
gggttggtta cgttgactcc ccgtaggtt tggtttaaat atgatgaagt ggacggaagg 5340
aaggaggaag acaaggaagg ataaggttgc aggcctgtg caaggtaaga agatggaat 5400
ttgatagagg tacgctacta tacttatact atacgctaag ggaatgcttg tatttatacc 5460
ctataacccc taataacccc ttatcaattt aagaaataat ccgcataagc ccccgcttaa 5520
aaattggtat cagagccatg aataggctta tgacaaaaac tcaagaggat aaaacctcac 5580
caaaatacga aagagttctt aactctaaag ataaaagatc tttcaagatc aaaactagtt 5640
ccctcacacc ggtgacgggg atcgcgatgg gtac 5674

```

&lt;210&gt; 56

&lt;211&gt; 6046

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz: binary  
vector pSUN1

&lt;400&gt; 56

```

ttccatggac atacaaatgg acgaacggat aaaccttttc acgccctttt aaatatccga 60

```

ttatttcta	aaacgctctt	ttctcttagg	tttaccgcgc	aatatctcct	gtcaaacact	120
gatagtttaa	actgaaggcg	ggaaacgaca	atcagatcta	gtaggaaaca	gctatgacca	180
tgattacgcc	aagcttgcat	gcctgcaggt	cgactctaga	ctagtggatc	cgatatcgcc	240
cgggctcgag	gtaccgagct	cgaattcact	ggcgcgcgtt	ttacaacgac	tcagctgctt	300
ggtaataatt	gtcattagat	tgtttttatg	catagatgca	ctcgaaatca	gccaatttta	360
gacaagtatc	aaacggatgt	taattcagta	cattaaagac	gtccgcaatg	tgttattaag	420
ttgtctaagc	gtcaatttgt	ttacaccaca	atatatcctg	ccaccagcca	gccaacagct	480
ccccgaccgg	cagctcggca	caaaatcacc	acgcgttacc	accacgccgg	ccggccgcat	540
ggtgttgacc	gtgttcgccc	gcattgccga	gttcgagcgt	tccctaatac	tcgaccgcac	600
ccggagcggg	cgcgaggccg	ccaaggcccg	aggcgtgaag	tttggccccc	gccctaccct	660
caccccgcca	cagatcgccg	acgcccgcga	gctgatcgac	cagggaaggcc	gcaccgtgaa	720
agaggcggct	gcactgcttg	gcgtgcatcg	ctcgaccctg	taccgcgcac	ttgagcgcag	780
cgaggaagtg	acgcccaccg	aggccaggcg	gcgcggtgcc	ttccgtgagg	acgcattgac	840
cgaggccgac	gccctggcgg	ccgccgagaa	tgaacgcaa	gaggaacaag	catgaaaccg	900
caccaggacg	gccaggacga	accgtttttc	attaccgaag	agatcgaggc	ggagatgac	960
gcggccgggt	acgtgttcga	gccgcccgcg	cacgtctcaa	ccgtgcggct	gcatgaaatc	1020
ctggccgggt	tgtctgatgc	caagctggcg	gcctggccgg	ccagcttgcc	cgctgaagaa	1080
accgagcgcc	gccgtctaaa	aaggctgagt	gtatttgagt	aaaacagctt	gcgtcatgcg	1140
gtcgctcgct	atatgatcgc	atgagtaaat	aaacaaatac	gcaaggggaa	cgcatgaagg	1200
ttatcgctgt	acttaaccag	aaaggcgggt	caggcaagac	gaccatcgca	acccatctag	1260
cccgcgccct	gcaactcgcc	ggggccgatg	ttctgttagt	cgattccgat	ccccaggcca	1320
gtgcccgcga	ttgggcggcc	gtgcccgaag	atcaaccgct	aaccgttgct	ggcatcgacc	1380
gcccagcgat	tgaccgcgac	gtgaaggcca	tcggccggcg	cgacttcgta	gtgatcgacg	1440
gagcgcccca	ggcggcggac	ttggctgtgt	ccgcgatcaa	ggcagccgac	ttcgtgctga	1500
ttccggtgca	gccaagccct	tacgacatat	gggccaccgc	cgacctgggt	gagctgggta	1560
agcagcgcat	tgaggtcacg	gatggaaggc	tacaagcggc	ctttgtcgtg	tcgcgggcga	1620
tcaaaggcac	gcgcacgccc	ggtgaggttg	ccgaggcgct	ggccgggtac	gagctgcccc	1680
ttcttgagtc	ccgtatcacg	cagcgcgtga	gctaccagg	cactgccgcc	gccggcaca	1740
ccgttcttga	atcagaacct	gagggcgacg	ctgcccgcga	ggtccaggcg	ctggccgctg	1800
aaattaaatc	aaaactcatt	tgagttaatg	aggtaaagag	aaaatgagca	aaagcacaaa	1860
cacgctaagt	gccggccgct	cgagcgcacg	cgacagcaag	gctgcaacgt	tggccagcct	1920
ggcagacacg	ccagccatga	agcgggtcaa	ctttcagttg	ccggcggagg	atcacaccaa	1980
gctgaagatg	tacgcggtac	gccaaggcaa	gaccattacc	gagctgctat	ctgaatacat	2040
cgcgagccta	ccagagtaaa	tgagcaaatg	aataaatgag	tagatgaatt	ttagcggcta	2100
aaggaggcgg	catggaaaat	caagaacaac	caggcaccga	cgccgtggaa	tgccccatgt	2160
gtggaggaac	gggcgggttg	ccaggcgtaa	gcggctgggt	tgtctgccgg	ccctgcaatg	2220
gcaactggaac	ccccaaagcc	gaggaatcgg	cgtgagcggg	cgcaaaccat	ccggcccggt	2280
acaaatcggc	gcggcgctgg	gtgatgacct	ggtggagaag	ttgaaggccg	cgcaggccgc	2340
ccagcggcaa	cgcatcgagg	cagaagcacg	ccccggtgaa	tcgtggcaag	cggccgctga	2400
tcgaatccgc	aaagaatccc	ggcaaccgcc	ggcagccggt	gcgcgctcga	ttaggaagcc	2460
gccaaggggc	gacgagcaac	cagatttttt	cgttcgcgat	ctctatgacg	tgggcacccg	2520
cgatagtcgc	agcatcatgg	acgtggccgt	tttccgtctg	tcgaagcgtg	accgacgagc	2580
tggcgaggtg	atccgctacg	agcttccaga	cgggcacgta	gaggtttccg	cagggccggc	2640
cggcatggcc	agtgtgtggg	attacgacct	ggtactgatg	gcgggtttcc	atctaaccga	2700
atccatgaac	cgataccggg	aagggaaggg	agacaagccc	ggccgcgtgt	tccgtccaca	2760
cgttgccggac	gtactcaagt	tctgccggcg	agccgatggc	ggaaagcaga	aagacgacct	2820
ggtagaaacc	tgcatctcgt	taaacaccac	gcacgttgcc	atgcagcgta	cgaagaaggc	2880
caagaacggc	cgccgtgtga	cggatatccga	gggtgaagcc	ttgattagcc	gctacaagat	2940
cgtaaagagc	gaaaccgggc	ggccggagta	catcgagatc	gagctagctg	attggatgta	3000
ccgcgagatc	acagaaggca	agaaccggga	cgtgctgacg	gttcaccccc	attacttttt	3060
gatcgatccc	ggcatcggcc	gtttttctta	ccgcctggca	cgccgcgccg	caggcaaggc	3120
agaagccaga	tggttgttca	agacgatcta	cgaacgcagt	ggcagcgccg	gagagttcaa	3180
gaagttctgt	ttcaccgtgc	gcaagctgat	cgggtcaa	gacctgccgg	agtacgattt	3240
gaaggaggag	gcggggcagg	ctggcccgat	cctagtcatg	cgctaccgca	acctgatcga	3300
gggcaagca	tccgcgggtt	cctaattgtac	ggagcagatg	ctagggcaaa	ttgccctagc	3360
aggggaaaaa	ggtcgaaaag	gtctctttcc	ttgggatagc	acgtacattg	ggaacccaaa	3420
gccgtacatt	gggaaccgga	accggtacat	tgggaaccca	aagccgtaca	ttgggaaccg	3480

```

gtcacacatg taagtgactg atataaaaaga gaaaaaaggc gattttttccg cctaaaaactc 3540
tttaaaactt attaaaaactc ttaaaaccgg cctggcctgt gcataactgt ctggccagcg 3600
cacagccgaa gagctgcaaa aagcgcctac ccttcggctg ctgcgctccc tacgccccgc 3660
cgcttcgctg cggcctatcg cggccgctgg ccgctcaaaa atggctggcc tacggccagg 3720
caatctacca gggcgcggac aagccgcgcc gtcgccactc gaccgccggc gccacatca 3780
aggcacctg cctcgcgcgt ttcggtgatg acggtgaaaa cctctgacac atgcagctcc 3840
cgggagacggg cacagcttgt ctgtaagcgg atgccgggag cagacaaggc cgtcagggcg 3900
cgtcagcggg tgttggcggg tgtcggggcg cagccatgac ccagtcacgt agcgatagcg 3960
gagtgtatac tggcttaact atgcggcatc agagcagatt gtactgagag tgcaccatc 4020
gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac cgcatcaggc gctcttccgc 4080
ttcctcgctc actgactcgc tgcgctcggg cgttcggctg cggcgagcgg tatcagctca 4140
ctcaaaggcg gtaatacggg atccacaga atcaggggat aacgcaggaa agaactgtg 4200
tagcaaaagg cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca 4260
agggctccgc cccctgacg agcatcacia aaatcgacgc tcaagtcaga ggtggcgaaa 4320
cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc 4380
tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc 4440
gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcggtc gctccaagct 4500
gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg 4560
tcttgagtc aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag 4620
gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggg gccctaacta 4680
cggctacact agaaggacag tatttgggat ctgcgctctg ctgaagccag ttaccttcgg 4740
aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtgtttttt 4800
tgtttgcaag cagcagatta cgcgcagaaa aaaggatct caagaagatc ctttgatctt 4860
ttctacgggg tctgacgctc agtggaaacg aaactcacgt taagggattt tggatcatgca 4920
tgatatatct cccaatttgt gtagggttta ttatgcacgc ttaaaaaataa taaaagcaga 4980
cttgacctga tagtttggct gtgagcaatt atgtgcttag tgcattctaac gcttgagtta 5040
agcgcgcgcg cgaagcggcg tcggcttgaa cgaatttcta gctagacatt atttgccgac 5100
taccttggtg atctcgcctt tcacgtagtg gacaaattct tccaactgat ctgcgcgcga 5160
ggccaagcga tcttcttctt gtccaagata agcctgtcta gcttcaagta tgacgggctg 5220
atactgggcc ggcaggcgtt ccattgcccc gtcggcagcg acatccttcg gcgcgatttt 5280
gccggttact gcgctgtacc aaatgcggga caacgtaagc actacatttc gctcatcgcc 5340
agcccagtcg ggcggcgagt tccatagcgt taaggtttca tttagcgctt caaatagatc 5400
ctgttcagga accggatcaa agagttcctc cgccgctgga cctaccaagg caacgctatg 5460
ttctcttgct ttgtcagca agatagccag atcaatgtcg atcgtggctg gctcgaagat 5520
acctgcaaga atgtcattgc gctgccatto tccaaattgc agttcgcgct tagctggata 5580
acgccacgga atgatgtcgt cgtgcacaa aatggtgact tctacagcgc ggagaatctc 5640
gctctctcca ggggaagccg aagtttccaa aaggtcgttg atcaaagctc gccgcgttgt 5700
ttcatcaagc cttacgggtc ccgtaaccag caaatcaata tcaactgtgtg gcttcaggcc 5760
gccatccact gcggagccgt acaaatgtac ggccagcaac gtcggttcga gatggcgctc 5820
gatgacgcca actacctctg atagttgagt cgatacttcg gcgatcaccg cttcccccat 5880
gatgtttaac tttgttttag ggcgactgcc ctgctgcgta acatcggttc tgctccataa 5940
catcaaacat cgacccacgg cgtaacgcgc ttgctgcttg gatgcccgag gcatagactg 6000
taccctcaaaa aaacagtcac aacaagccat gaaaaccgcc actgcg 6046

```

&lt;210&gt; 57

&lt;211&gt; 9838

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz: Transgenic  
expression vector for codA dsRNA pSUN1-codA-RNAi

&lt;400&gt; 57

```

cgaattcact ggccgctggt ttacaacgac tcagctgctt ggtaataatt gtcattagat 60
tgtttttatg catagatgca ctcgaaatca gccaatttta gacaagtatc aaacggatgt 120
taattcagta cattaaagac gtccgcaatg tgttattaag ttgtctaagc gtcaatttgt 180
ttacaccaca atatatctg ccaccagcca gccaacagct ccccgaccgg cagctcggca 240
caaaatcacc acgcgttacc accacgcggc ccggccgcat ggtgttgacc gtgttcgccc 300

```

gcattgccga	gttcgagcgt	tcctaatca	tcgaccgcac	ccggagcggg	cgcgaggccg	360
ccaaggcccc	aggcgtgaag	tttggcccc	gccctaccct	caccccgga	cagatcgcg	420
acgcccgcga	gctgatcgac	caggaaggcc	gcaccgtgaa	agaggcggct	gactgcttg	480
gcgtgcatcg	ctcgaccctg	taccgcgcac	ttgagcgcag	cgaggaagtg	acgcccaccg	540
aggccaggcg	gcgcggtgcc	ttccgtgagg	acgcattgac	cgaggccgac	gccctggcgg	600
ccgccgagaa	tgaacgcaa	gaggaacaag	catgaaaccg	caccaggacg	gccaggacga	660
accgtttttc	attaccgaag	agatcgaggc	ggagatgata	gcggccgggt	acgtgttcga	720
gccgcccgcg	cacgtctcaa	ccgtgcggct	gcatgaaatc	ctggccgggt	tgtctgatgc	780
caagctggcg	gcctggccgg	ccagcttggc	cgctgaagaa	accgagcgcc	gccgtctaaa	840
aaggtgatgt	gtatttgagt	aaaacagctt	gcgtcatgcg	gtcgctgcgt	atatgatgcg	900
atgagtaaat	aaacaaatac	gcaaggggaa	cgcatgaagg	ttatcgctgt	acttaaccag	960
aaaggcgggt	caggcaagac	gaccatcgca	acccatctag	cccgcgccct	gcaactcgcc	1020
ggggccgatg	ttctgttagt	cgattccgat	ccccagggca	gtgcccgcga	ttgggcccgc	1080
gtgcgggaag	atcaaccgct	aaccgttgtc	ggcatcgacc	gcccgcacat	tgaccgcgac	1140
gtgaaggcca	tcggccggcg	cgacttcgta	gtgatcgacg	gagcgcccca	ggcgccggac	1200
ttggctgtgt	ccgcgatcaa	ggcagccgac	ttcgtgctga	ttccgggtga	gccaagccct	1260
tacgacatat	gggccaccgc	cgacctggtg	gagctgggta	agcagcgcat	tgaggtcacg	1320
gatggaaggc	tacaagcggc	ctttgtcgtg	tcgcgggcga	tcaaaggcac	gcgcacggc	1380
ggtgaggttg	ccgagggcgt	ggccgggtac	gagctgcca	ttcttgagtc	ccgtatcacg	1440
cagcgctga	gctaccacgg	cactgccgcc	gccggcaca	ccgttcttga	atcagaacct	1500
gagggcgacg	ctgcccgcga	ggtccaggcg	ctggccgctg	aaattaaatc	aaaactcatt	1560
tgagttaatg	aggtaaagag	aaaatgagca	aaagcaca	cacgctaagt	gccggccgct	1620
cgagcgacg	cagcagcaag	gctgcaacgt	tggccagcct	ggcagacacg	ccagccatga	1680
agcgggtcaa	ctttcagttg	ccggcggagg	atcacacca	gctgaagatg	tacgcggtac	1740
gccaaggcaa	gaccattacc	gagctgctat	ctgaatacat	cgcgacagta	ccagagtaaa	1800
tgagcaaatg	aataaatgag	tagatgaatt	ttagcggcta	aaggaggcgg	catggaaaat	1860
caagaacaac	caggcaccga	cgccgtggaa	tgccccatgt	gtggagggaac	gggcccgttg	1920
ccaggcgtaa	gcggctgggt	tgtctgccgg	ccctgcaatg	gactggaac	cccaagccc	1980
gaggaatcgg	cgtgagcgg	cgcaaaccat	ccggcccgg	acaaatcggc	gcggcgctgg	2040
gtgatgacct	ggtggagaag	ttgaaggccg	cgcagccgcg	ccagcggcaa	cgcatcgagg	2100
cagaagcacg	ccccggtgaa	tcgtggcaag	cggccgctga	tcgaatccgc	aaagaatccc	2160
ggcaaccgcc	ggcagccgg	gcgcgctcga	ttaggaagcc	gccaaggggc	gacgagcaac	2220
cagatTTTTT	cgttccgatg	ctctatgacg	tgggcacccg	cgatagtcgc	agcatcatgg	2280
acgtggccgt	tttccgtctg	tcgaagcgtg	accgacgagc	tggcgagggtg	atccgctacg	2340
agcttccaga	cgggcacgta	gaggtttccg	cagggccggc	cggcattggcc	agtgtgtggg	2400
attacgacct	ggtactgatg	gcggtttccc	atctaaccga	atccatgaac	cgataccggg	2460
aagggaaggg	agacaagccc	ggccgcgtgt	tccgtccaca	cgttgcggac	gtactcaagt	2520
tctgccggcg	agccgatggc	ggaaagcaga	aagacgacct	ggtagaaacc	tgcatccggt	2580
taaacaccac	gcacgttgcc	atgcagcgta	cgaagaaggc	caagaacggc	cgccgtggtga	2640
cggatatccga	gggtgaagcc	ttgattagcc	gctacaagat	cgtaaagagc	gaaaccgggc	2700
ggccggagta	catcgagatc	gagctagctg	attggatgta	ccgcgagatc	acagaaggca	2760
agaacccgga	cgtgctgacg	gttcaccccg	attacttttt	gatcgatccc	ggcatcggcc	2820
gttttctcta	ccgcctggca	cgccgcgcgc	caggcaaggc	agaagccaga	tggttgttca	2880
agacgatcta	cgaacgcagt	ggcagcgccg	gagagttaa	gaagtctctg	ttcacctgtc	2940
gcaagctgat	cgggtcaaat	gacctgccgg	agtagatttt	gaaggaggag	gcggggcagg	3000
ctggcccgat	cctagtcatg	cgctaccgca	acctgatcga	gggcgaagca	tccgcgggtt	3060
cctaattgtac	ggagcagatg	ctagggcaaa	ttgccctagc	aggggaaaaa	ggtcgaaaag	3120
gtctctttcc	tgtggatagc	acgtacattg	ggaacccaaa	gccgtacatt	gggaaccgga	3180
accggtacat	tgggaaccca	aagccgtaca	ttgggaaccg	gtcacacatg	taagtgactg	3240
atataaaaga	gaaaaaaggc	gatttttccg	cctaaaactc	tttaaaactt	attaaaactc	3300
ttaaaaccgg	cctggcctgt	gcataactgt	ctggccagcg	cacagccgaa	gagctgcaaa	3360
aagcgcctac	ccttcgggtcg	ctgcgctccc	tacgccccgc	cgcttcgcgt	cggcctatcg	3420
cgcccgctgg	ccgctcaaaa	atggctggcc	tacggccagg	caatctacca	gggcgcggac	3480
aagccgcgcc	gtcgcacttc	gaccgcgggc	gcccacatca	aggcaccctg	cctcgcgctg	3540
ttcgggtgatg	acggtgaaaa	cctctgacac	atgcagctcc	cgagagacgg	cacagcttgt	3600
ctgtaagcgg	atgccgggag	cagacaagcc	cgtcaggggc	cgtcagcggg	tgttggcggg	3660
tgtcggggcg	cagccatgac	ccagtcacgt	agcgatagcg	gagtgataac	tggtttaact	3720

atgcgggcatc	agagcagatt	gtactgagag	tgcaccatat	gcgggtgtgaa	ataccgcaca	3780
gatgcgtaag	gagaaaaatac	cgcatcaggc	gctcttccgc	ttcctcgctc	actgactcgc	3840
tgcgctcggg	cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	3900
tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	3960
ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	4020
agcatcacaa	aaatcgacgc	tcaagtcaga	ggtggcgaaa	cccgcacagga	ctataaagat	4080
accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	tggtccgacc	ctgccgctta	4140
ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	4200
gtaggtatct	cagttcgggtg	taggtcgttc	gctccaagct	gggtgtgtg	cacgaacccc	4260
ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtc	aaccgggtaa	4320
gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtag	4380
taggcgggtg	tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	4440
tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	4500
gatccggcaa	acaaaccacc	gctggtagcg	gtggtttttt	tggttgcaag	cagcagatta	4560
cgcgagaaa	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	4620
agtggaaacga	aaactcacgt	taagggatct	tggtcatgca	tgatatactt	cccaatttgt	4680
gtagggctta	ttatgcacgc	ttaaaaataa	taaaagcaga	cttgacctga	tagtttggtc	4740
gtgagcaatt	atgtgcttag	tgcatctaac	gcttgagtta	agccgcgcgc	cgaagcggcg	4800
tcggcttgaa	cgaatttcta	gctagacatt	atttgccgac	taccttggtg	atctcgcctt	4860
tcacgtagt	gacaaattct	tccaactgat	ctgcgcgcga	ggccaagcga	tcttcttctt	4920
gtccaagata	agcctgtcta	gcttcaagta	tgacgggctg	atactggggc	ggcaggcgct	4980
ccattgccca	gtcggcagcg	acatccttcg	gcgcgatttt	gccggttact	gcgctgtacc	5040
aatgcgggga	caacgtaagc	actacatttc	gctcatcgcc	agcccagtcg	ggcggcgagt	5100
tccatagcgt	taaggtttca	tttagcgctt	caaataagatc	ctgttcagga	accggatcaa	5160
agagttcctc	cgccgctgga	cctaccaagg	caacgctatg	ttctcttgct	tttgtagca	5220
agatagccag	atcaatgtcg	atcgtggctg	gtcgaagat	acctgcaaga	atgtcattgc	5280
gctgccattc	tccaaattgc	agttcgcgct	tagctggata	acgccacgga	atgatgtcgt	5340
cgtgcacaac	aatggtgact	tctacagcgc	ggagaatctc	gctctctcca	ggggaagccg	5400
aagttttcaa	aaggtcgttg	atcaaagctc	gccgcgttgt	ttcatcaagc	cttacggtca	5460
ccgtaaccag	caaatacaata	tactgtgtg	gcttcaggcc	gccatccact	gcggagccgt	5520
acaaatgtac	ggccagcaac	gtcggttcga	gatggcgctc	gatgacgcca	actacctctg	5580
atagttgagt	cgatacttcg	gcgatcaccg	cttcccccat	gatgtttaac	tttggttttag	5640
ggcgactgcc	ctgctgcgta	acatcggttg	tgctccataa	catcaaacat	cgacccacgg	5700
cgtaacgcgc	ttgctgcttg	gatgcccag	gcatagactg	tacccccaaa	aaacagtcac	5760
aacaagccat	gaaaaccgcc	actgcgttcc	atggacatac	aaatggacga	acggataaac	5820
cttttcacgc	ccttttaaat	atccgattat	tctaataaac	gctcttttct	cttaggttta	5880
cccgccaata	tatcctgtca	aacactgata	gtttaaactg	aaggcgggaa	acgacaatca	5940
gatctagtag	gaaacagcta	tgaccatgat	tacgccaagc	ttgcatgcct	gcaggctcgac	6000
tctagactag	tggatccgat	atcgcccggg	ctcgaggtag	ccatcgcgat	ccccgtcacc	6060
ggtgtgaggg	aactagtttt	gatcttgaaa	gatcttttat	cttttagagtt	aagaactctt	6120
tcgtattttg	gtgaggtttt	atcctcttga	gttttggtca	tagacctatt	catggctctg	6180
ataccaattt	ttaagcgggg	gcttatgcgg	attatttctt	aaattgataa	ggggttatta	6240
gggggtatag	ggtataaata	caagcattcc	cttagcgtat	agtataagta	tagtagcgta	6300
cctctatcaa	atttccatct	tcttaccttg	cacagggcct	gcaaccttat	ccttccttgt	6360
cttctctctt	ccttccgctc	acttcatcat	atttaaacca	aacctacggg	ggagtcaacg	6420
taaccaaccc	tgcttagca	tcttttccct	aacggcctcc	tgcttaagcg	gtacttctag	6480
cttcgaacgg	cgtctgggct	ccagggttag	tcgtctcgtg	tctggtttat	attcacgaca	6540
aagatctata	gggacttttag	gagatctgga	ttttagtact	ggatttttgt	tttaggaatt	6600
agaaatttta	ttgatagaag	tattttacaa	atacaaatac	atactaaggg	tttcttatat	6660
gctcaacata	tgagcgaaac	cctataagaa	ccctaatctc	cccttatcgg	gaaactactc	6720
acacattatt	tatggagaaa	atagagagag	atagatttgt	agagagagac	tggtgatttc	6780
agcgtaccga	attcggctaa	cagtgtcgaa	taacgcttta	caaacaatta	ttaacgccc	6840
gttaccaggc	gaagaggggc	tgtggcagat	tcactctgcag	gacggaaaaa	tcagcgccat	6900
tgatgcgcaa	tccggcggtga	tgcccataac	tgaaaacagc	ctggatgccc	aacaagggtt	6960
agttataccg	ccgtttgtgg	agccacatat	tcacctggac	accacgcaaa	ccgcccggaca	7020
accgaactgg	aatcagtcgg	gcacgctgtt	tgaaggcatt	gaacgctggg	ccgagcgcaa	7080
agcgttatta	acccatgacg	atgtgaaaca	acgcgcatgg	caaacgctga	aatggcagat	7140



```

tgccaacggc attcagcatg tgcgtaccca tgtcgaatgt tgggatgcaa cgctaactgc 7200
gctgaaagca atgctggaag tgaagcagga agtcgcgcgc tggattgac tgcaaatcgt 7260
cgcttccct caggaaggga ttttgtcggg tccggtgata cctgcacatc aacaaatctt 7320
ggctatatat tagaaaagt ataaattaaa atatacacac ttataaacta cagaaaagca 7380
attgctatat actacattct tttattttga aaaaaatatt tgaaatatta tattactact 7440
aattaatgat aattattata tatatatcaa aggtagaagc agaaacttac gtagtcgacg 7500
acaaaatccc ttcctgaggg aaggcgacga tttgcagatc aatccacggc gcgacttcct 7560
gcttcacttc cagcattgct ttcagcgagc ttagcggtgc atccgaaaca tcgacatggg 7620
tacgcacatg ctgaatgccg ttggcaatct gccatttcag cgtttgccat gcgcgttgtt 7680
tcacatcgtc atgggttaat aacgctttgc gctcggccca gcgttcaatg ccttcaaaca 7740
gcgtgccgga ctgattccag ttcggttgtc cggcggtttg cgtggtgtcc aggtgaatat 7800
gtggctccac aaacggcggt ataactaaac cttgttcggc atccaggctg ttttcagtta 7860
tgggcatcac gccggattgc gcatcaatgg cgctgatttt tccgtcctgc agatgaatct 7920
gccacagccc ctcttcgcct ggtaaccggg cggttaataat tgtttgtaaa gcgttattcg 7980
acactgttag ccaagcttgc atgcctgcga gtcgagctct tgttttttac tttggttcat 8040
gacactcaga gacttgagag aagcaatata tagacttttt tttgtttttt tttgtggtc 8100
acgtttatct tcctattgga gacggtaacg aagatcgaa cgtgtgtgga aatgaaacaa 8160
ggtgggacta gccacgtgg tttcttttct ctgcattgat ttgtttttgt tttttttgta 8220
aagttcacat caaacctact aataattgag aagaaaaata aaatctattg attgattaaa 8280
ccagccgatg ctttatgtct gaataaaaa aagaagtga aaccccgttt aagaattaca 8340
acggtgggtt acaaagtatt tggacacaat aaatccaaac gaaataaaac aaaatggaga 8400
actaccaa ataaaaacaaa taaaaaactt aaaagaattt attccatttt ttttcccgta 8460
gaatttatct ttttatggat tccttaaatc catattttgat gcattttgat tcctcataat 8520
aggtaataat atatactatg ttatagatat gtttctaat cgtattaacc tacctttttt 8580
tggtcgtacg attctaccta ataatttga acggaattga tgttttggac cacttagaaa 8640
gtattttttt tttggtttgt cttagctgta tttcattaaa tataaattta aataagaaat 8700
gtcataaata aaatttgacg tatagatttt ttaaattccat tttatgttat ttaattttg 8760
aatgtgagtg ttggctccta tttaatctta ggatgggtta atactaagtt ttccttaatg 8820
aattatctca gagaaactgg attaaataaa ctaaaaaata gatcaatgtg ttttgggtccg 8880
gtcaaatact tttggattta ctattattgg cgaaaagaaa gtctcatata gtaaatcata 8940
ttcctacaag agaaatcaaa atttttgaat taacatggat tgtatagttt cttatataac 9000
caattagttc gcatcaagaa aaccaaacc caattaataa tcaaacgggc ttggtaggaa 9060
tatttcattg cagctttcag ataaaagaaa aaaacacaca ctcaagtctt ttatttcac 9120
tttcttactt gcaggaactc aaattccact ttgccacttt tctttacaaa taacacaaa 9180
ttgtcaatga aacgaaatag tctttttatg caaacactgt ttgtctttt tcgacacgt 9240
ttctgattgt gacagccatc catatatata gggaaatgta aacaacaaca tgtgaagtca 9300
catatacgta atggttttag atagcttcta ttttcgttgt caatattagt cattccaaa 9360
catttttaag aaaaataaat taatatatgt atattcttgg aactaatgta tgtggaaata 9420
cagtaactta attattaaac attctaaatg caaatatgca aagaaaaaaa agaaaagaac 9480
acaactgaaa tcaaagccag attcataata attggctaca tgggtgtaga atgtagggt 9540
acacaacatc cagaattgaa cactcaaatt ggatgataga tggataatct ttagatacaa 9600
gagaattggg tctcttccat tattaacgaa aataaagaaa aaaagttag cataaaagt 9660
tgaaactcaa cataacattt tgaacttgac tccttcatag gagtgacatg aactgacgaa 9720
tcacaaccga ttacttggtt gagtcattct ccgctttctc caccttcgaa atgaatgtga 9780
ccggtttctt cgggtgctca tttacggtca agtgtaaaac atctggtctc gacgagct 9838

```

&lt;210&gt; 58

&lt;211&gt; 14184

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz: Expression  
vector pSUN1-codA-RNAi-At.Act.-2-At.Als-R-ocST

&lt;400&gt; 58

```

ctgcttggtg ataattgtca ttagattgtt tttatgcata gatgcactcg aaatcagcca 60
attttagaca agtatcaaac ggatgttaat tcagtacatt aaagacgtcc gcaatgtgtt 120
attaagttgt ctaagcgtca atttgtttac accacaatat atcctgccac cagccagcca 180

```

acagctcccc gaccggcagc tcggcacaaa atcaccacgc gttaccacca cgccggccgg 240  
 ccgcatgggtg ttgaccgtgt tcgcccggcat tgcgagttc gagcgttccc taatcatcga 300  
 ccgcacccgg agcgggcgcg aggcgcgcaa ggcccaggc gtgaagtttg gccccgcgc 360  
 taccctcacc ccggcacaga tcgcgacgc ccgagagctg atcgaccagg aaggccgcac 420  
 cgtgaaagag gcggctgcac tgcttggcgt gcacgctcg accctgtacc gcgcacttga 480  
 gcgcagcgag gaagtgcgc ccaccgaggg caggcggcgc ggtgccttcc gtgaggacgc 540  
 attgaccgag gccgacgccc tggcggccgc cgagaatgaa cgccaagagg aacaagcatg 600  
 aaaccgcacc aggacggcca ggacgaaccg tttttcatta ccgaagagat cgaggcggag 660  
 atgatcgcg ccgggtacgt gttcgagccg cccgcgcacg tctcaaccgt gcggctgcat 720  
 gaaatcctgg ccggtttgtc tgatgccaag ctggcggcct ggccggccag cttggccgct 780  
 gaagaaaccg agcgccgccc tctaaaaagg tgatgtgtat ttgagtaaaa cagcttgctg 840  
 catgcggtcg ctgcgtatat gatgcgatga gtaataaac aaatacgcaa ggggaacgca 900  
 tgaaggttat cgctgtactt aaccagaaag gcgggtcagg caagacgacc atcgcaaccc 960  
 atctagcccc cgccctgcaa ctgcggggg ccgagttct gttagtcat tccgatcccc 1020  
 agggcagtg ccgcgattgg gcggccgtgc gggaagatca accgctaacc gttgtcggca 1080  
 tcgaccgccc gacgattgac cgcgacgtga aggccatcg ccggcgcgac ttcgtagtga 1140  
 tcgacggagc gcccaggcg gcggacttgg ctgtgtccgc gatcaaggca gccgacttcg 1200  
 tgctgattcc ggtgcagcca agcccttacg acatatgggc caccgcccac ctggtggagc 1260  
 tggtaaagca gcgcattgag gtcacggatg gaaggctaca agcggcctt gtcgtgtcgc 1320  
 gggcgatcaa aggcacgcgc atcggcgggtg aggttgccga ggcgctggcc ggggtacgagc 1380  
 tgccattct tgagtcccg atcacgcagc gcgtgagcta cccaggcact gccgcccgcg 1440  
 gcacaaccgt tcttgaatca gaaccgagg gcgacgctgc ccgcgaggtc caggcgtg 1500  
 ccgctgaaat taaatcaaaa ctcatcttag ttaatgaggt aaagagaaaa tgagcaaaag 1560  
 cacaacacg ctaagtgcgc gccgtccgag cgcacgcagc agcaaggctg caacgttggc 1620  
 cagcctggca gacacgccag ccatgaagcg ggtcaacttt cagttgcccg cgaggatca 1680  
 caccaagctg aagatgtacg cggtagcga aggcaagacc attaccgagc tgctatctga 1740  
 atacatcgcg cagctaccag agtaaatgag caaatgaata aatgagtaga tgaattttag 1800  
 cggctaaagg aggcggcatg gaaaatcaag aacaaccagg caccgacgcc gtggaatgcc 1860  
 ccatgtgtgg aggaacgggc ggttggccag gcgtaagcgg ctgggttgtc tgccggccct 1920  
 gcaatggcac tggaaacccc aagcccgagg aatcggcgtg agcggtcgca aaccatccg 1980  
 cccggtacaa atcggcggcg cgctgggtga tgacctggtg gagaagtga aggccgcgca 2040  
 ggccgcccag cggcaacgca tcgaggcaga agcacgcccc ggtgaatcgt ggcaagcggc 2100  
 cgctgatcga atccgcaaag aatcccggca acccgccgga gccggtgctc cgtcgattag 2160  
 gaagccgccc aagggcgacg agcaaccaga ttttttcgtt ccgatgctct atgacgtggg 2220  
 caccgcgat agtcgcagca tcatggacgt ggccgttttc cgtctgtcga agcgtgaccg 2280  
 acgagctggc gaggtgatcc gctacgagct tccagacggg cacgtagagg tttccgcagg 2340  
 gccggccggc atggccagtg tgtgggatta cgacctggta ctgatggcg tttcccatct 2400  
 aaccgaatcc atgaaccgat accgggaagg gaaggagagc aagcccggcc gcgtgttccg 2460  
 tccacacgtt gcggacgtac tcaagttctg ccggcgagcc gatggcggaa agcagaaaga 2520  
 cgacctggtg gaaacctgca ttcggttaaa caccacgcac gttgccatgc agcgtacgaa 2580  
 gaaggccaag aacggccgccc tgggtgacggt atccgagggg gaagccttga ttagccgcta 2640  
 caagatcgta aagagcgaaa ccgggcggcc ggagtacatc gagatcgagc tagctgattg 2700  
 gatgtaccgc gagatcacag aaggcaagaa cccggacgtg ctgacggttc accccgatta 2760  
 ctttttgatc gatcccggca tcggccggtt tctctaccgc ctggcacgcc gcgccgagc 2820  
 caaggcagaa gccagatggt tgttcaagac gatctacgaa cgcagtggca gcgccggaga 2880  
 gttcaagaag ttctgtttca ccgtgcgcaa gctgatcggg tcaaatgacc tgccggagta 2940  
 cgatttgaag gaggagggcg ggcaggctgg cccgatccta gtcatgcgct accgcaacct 3000  
 gatcgagggc gaagcatccg ccggttccta atgtacggag cagatgctag ggcaaatg 3060  
 cctagcaggg gaaaaaggtc gaaaaggtct ctttctgtg gatagcacgt acattgggaa 3120  
 cccaaagccg tacattggga accggaacct gtacattggg aacccaaagc cgtacattgg 3180  
 gaaccgggtc cacatgtaag tgactgatat aaaagagaaa aaaggcgatt tttccgcta 3240  
 aaactcttta aaacttatta aaactcttaa aaccgcctg gcctgtgcat aactgtctg 3300  
 ccagcgaca gccgaagagc tgcaaaaagc gcctaccctt cggtcgctgc gctccctacg 3360  
 ccccgccgct tcgctcgcc ctatcgccgc cgctggccgc tcaaaaatgg ctggccctacg 3420  
 gccaggcaat ctaccagggc gcggacaagc cgcgcgctcg cactcgacc gccggcgccc 3480  
 acatcaaggc accctgcctc gcgcgtttcg gtgatgacgg tgaanaacct tgacacatgc 3540  
 agctcccggg gacggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 3600

agggcgcgctc	agcgggtggtt	ggcgggtgtc	ggggcgcgagc	catgaccag	tcacgtagcg	3660
atagcggagt	gtatactggc	ttaactatgc	ggcatcagag	cagattgtac	tgagagtgc	3720
ccatatgcgg	tgtgaaatac	cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	3780
ttccgcttcc	tcgctcactg	actcgtctgc	ctcggctcgtt	cggctgcggc	gagcggatc	3840
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataacg	caggaaagaa	3900
catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	3960
tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	4020
gcgaaacccg	acaggactat	aaagatacca	ggcgtttccc	cctggaagct	ccctcggtcg	4080
ctctcctggt	ccgacctgc	cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	4140
cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	tcgggtgtagg	tcggtcgctc	4200
caagctgggc	tgtgtgcacg	aaccccccg	tcagcccgac	cgtgcgcct	tatccggtaa	4260
ctatcgctct	gagtccaacc	cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	4320
taacaggatt	agcagagcga	ggtatgtagg	cgggtgtaca	gagttcttga	agtgggtggc	4380
taactacggc	tacactagaa	ggacagtatt	tggatctctg	gctctgctga	agccagttac	4440
cttcggaaaa	agagttggta	gctcttgatc	cggcaaaaa	accaccgctg	gtagcggttg	4500
tttttttgtt	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	4560
gatcttttct	acggggtctg	acgctcagtg	gaacgaaaac	tcacgttaag	ggattttgggt	4620
catgcatgat	atatgtccca	atttgtgtag	ggcttattat	gcacgcttaa	aaataataaa	4680
agcagacttg	acctgatagt	ttggctgtga	gcaattatgt	gcttagtgca	tctaacgctt	4740
gagttaagcc	gcgcgcgcaa	ggggcgctcg	cttgaacgaa	tttctagcta	gacattattt	4800
gccgactacc	ttggtgatct	cgcctttcac	gtagtggaca	aattcttcca	actgatctgc	4860
gcgcgaggcc	aagcgatctt	cttcttgtcc	aagataagcc	tgtctagctt	caagtatgac	4920
gggctgatac	tgggcccggca	ggcgctccat	tgcccagtcg	gcagcgacat	ccttcggcgc	4980
gattttgccc	gttactgcgc	tgtaccaa	gcgggacaac	gtaagcacta	catttcgctc	5040
atcgccagcc	cagtcggggc	gcgagttcca	tagcgttaa	gtttcattta	gcgcctcaa	5100
tagatcctgt	tcaggaaccg	gatcaaagag	ttcctccgcc	gctggacct	ccaaggcaac	5160
gctatgttct	cttgcttttg	tcagcaagat	agccagatca	atgtcgatcg	tggctggctg	5220
gaagatacct	gcaagaatgt	cattgcgctg	ccattctcca	aattgcagtt	cgcgcttagc	5280
tgagataacg	cacggaatga	tgtcgtcgtg	cacaacaatg	gtgacttcta	cagcgaggag	5340
aatctcgctc	tctccagggg	aagccgaagt	ttccaaaagg	tcgttgatca	aagctcgccg	5400
cgttgtttca	tcaagcetta	cggtcaccgt	aaccagcaaa	tcaatatcac	tgtgtggctt	5460
caggccgcca	tccactgcgg	agccgtacaa	atgtacggcc	agcaacgtcg	gttcgagatg	5520
gcgctcgatg	acgccaaact	cctctgatag	ttgagtcgat	acttcggcga	tcaccgcttc	5580
ccccatgatg	tttaactttg	ttttagggcg	actgccctgc	tgcgtaacat	cgttgctgct	5640
ccataacatc	aaacatcgac	ccacggcgta	acgcgcttgc	tgcttggtatg	cccaggcat	5700
agactgtacc	ccaaaaaac	agtcataaca	agccatgaaa	accgccactg	cgttccatgg	5760
acatacaaat	ggacgaacgg	ataaaccttt	tcagccctt	ttaaatatcc	gattattcta	5820
ataaacgctc	ttttctctta	ggtttaccgg	ccaatatatc	ctgtcaaaca	ctgatagttt	5880
aaactgaagg	cgggaaacga	caatcagatc	tagtaggaaa	cagctatgac	catgattacg	5940
ccaagcttgc	atgcctgcag	gtcgactcta	gactagtgg	tccgatatcg	cccgggctcg	6000
aggtagccat	cgcgatcccc	gtcaccgggtg	tgagggaact	agttttgatc	ttgaaagatc	6060
ttttatcttt	agagtttaaga	actctttcgt	attttgggtga	ggttttatcc	tcttgagttt	6120
ttggtcataga	cctattcatg	gctctgatac	caatttttaa	gcgggggctt	atgcggatta	6180
tttcttaaat	tgataagggg	ttattagggg	gtatagggta	taaatacaag	cattccctta	6240
gcgtatagta	taagtatagt	agcgtacctc	tatcaaattt	ccatcttctt	accttgacac	6300
gggcctgcaa	ccttatcctt	ccttgtcttc	ctccttccct	ccgtccactt	catcatattt	6360
aaaccaaacc	tacgggggag	tcaacgtaac	caacctgcc	ttagcatctt	ttccctaacc	6420
gcctcctgcc	taagcggtag	ttctagcttc	gaacggcgct	tgggctccag	gttttagtcgt	6480
ctcgtgtctg	gtttatattc	acgacaaaga	tctataggga	ctttaggaga	tctggatttt	6540
agtactggat	tttggtttta	ggaattagaa	attttattga	tagaagtatt	ttacaaatac	6600
aaatacatat	taagggtttc	ttatatgctc	aacacatgag	cgaaacccta	taagaaccct	6660
aatttccctt	atcgggaaac	tactcacaca	ttatttatgg	agaaaaataga	gagagataga	6720
ttttagtaga	gagactgggtg	atttcagcgt	accgaattcg	attttcggct	aacagtgctg	6780
aataacgctt	tacaaacaat	tattaacgcc	cggttaccag	gcgaagaggg	gctgtggcag	6840
attcatctgc	aggacggaaa	aatcagcgcc	attgatgcgc	aatccggcgt	gatgcccata	6900
actgaaaaca	gcctggatgc	cgaacaaggt	ttagttatac	cgccgtttgt	ggagccacat	6960
attcacctgg	acaccacgca	aaccgccgga	caaccgaact	ggaatcagtc	cgccacgctg	7020

tttgaaggca	ttgaacgctg	ggccgagcgc	aaagcggttat	taacccatga	cgatgtgaaa	7080
caacgcgcat	ggcaaacgct	gaaatggcag	attgccaacg	gcattcagca	tgtgcgtacc	7140
catgtcgcg	tttcggatgc	aacgctaact	gcgctgaaaag	caatgctgga	agtgaagcag	7200
gaagtcgcgc	cgtggattga	tctgcaaatc	gtcgccttcc	ctcaggaagg	gattttgtcg	7260
gatccggtga	tacctgcaca	tcaacaaatt	ttgggtcatat	attagaaaag	ttataaatta	7320
aaatatacac	acttataaac	tacagaaaag	caattgctat	atactacatt	ctttttatatt	7380
gaaaaaaata	tttgaaatat	tatatattacta	ctaattaatg	ataattatta	tatatatatc	7440
aaaggtagaa	gcagaaactt	acgtagtcga	cgacaaaatc	ccgtcctgag	ggaaggcgac	7500
gattttgcaga	tcaatccacg	gcgcgacttc	ctgcttcact	tccagcattg	cttttcagcgc	7560
agttagcgtt	gcacccgaaa	catcgacatg	ggtacgcaca	tgctgaatgc	cgttggcaat	7620
ctgccatttc	agcgtttgcc	atgcgcggtg	tttcacatcg	tcatgggtta	ataacgcttt	7680
gcgctcggcc	cagcgttcaa	tgccttcaaa	cagcgtgccg	gactgattcc	agttcgggtg	7740
tccggcgggt	tgctgtgtgt	ccaggtgaat	atgtggctcc	acaaacggcg	gtataactaa	7800
accttggtcg	gcacccaggc	tgttttcagt	tatgggcac	acgccggatt	gcgcacatca	7860
ggcgttgatt	tttcgcgtcct	gcagatgaat	ctgccacagc	ccctcttcgc	ctggtaaccg	7920
ggcggttaata	attggtttgta	aagcgttatt	cgacactgtt	agccaagctt	gcacgcctgc	7980
aggtcgactc	tagaggatcc	ccgatccact	cgagtctttg	ttttttactt	tggttcatga	8040
cactcagaga	cttgagagaa	gcaatatata	gacttttttt	tggttttttt	ttgtggtcac	8100
gtttattttt	ctattggaga	cggtaacgaa	gatcgaacct	gtgggtggaa	tgaaacmagg	8160
tgggactagc	ccacgtgggt	tcttttctct	gcattgattt	gtttttgttt	tttytgtaaa	8220
gttcacatca	aacctactaa	taattgagaa	gaaaaataaa	atctattgat	tgattaaacc	8280
agccgatgct	ttatgtctga	atataaaaaa	gaagtgaaaa	ccccgtttta	gaattacaac	8340
ggtggtttac	aaagtatttg	gacacaataa	atccaaacga	aataaaacaa	aatggagaac	8400
taccaataaa	aaaacaaata	aaaaacttaa	aagaatttat	tccatttttt	ttcccgtaga	8460
atttattctt	ttatggattc	cttaaatcca	tatttgatgc	attttgattc	ctcataatag	8520
gtaataatat	atactatgtt	atagatatgt	ttctaattcg	tattaacctt	cttttttttg	8580
gtcgtacgat	tctacctaat	aatattgaac	ggaattgatg	ttttggacca	cttagaaagt	8640
attttttttt	tggtttgtct	tagctgtatt	tcattaaata	taaatttaaa	taagaaatgt	8700
cataaataaa	atttgacgta	tagatttttt	aaatccattt	tatgttattt	aatatttgaa	8760
atgtgagttt	ggctcctatt	taatcttagg	atgggttaat	actaagtttt	ccttaatgaa	8820
ttatctcaga	gaaactggat	taaataaact	aaaaaataga	tcaatgtgtt	ttgggtccgg	8880
caaatatctt	tggatttact	attattggcg	aaaagaaaag	ctcatatagt	aaatcatatt	8940
cctacaagag	aaatcaaaat	ttttgaatta	acatggattg	tatagtttct	tatataacca	9000
attagttcgc	atcaagaaaa	ccaaacccca	attaataatc	aaacgggctt	ggttaggaata	9060
tttcattgca	gctttcagat	aaaagaaaaa	aacacacact	caagtctttt	atttcatctt	9120
tcttacttgc	aggaactcaa	attccacttt	gccacttttc	tttacaataa	aacacaaatt	9180
gtcaatgaaa	cgaaatagtc	tttttatgca	aacactgttt	gtcttttttc	gatcacgttt	9240
ctgattgtga	cagccatcca	tatatatagg	gaatgtaaaa	caacaacatg	tgaagtcaca	9300
tatacgtaat	ggtttagcat	agcttctatt	ttcgttgtca	atattagtca	ttccaaaaca	9360
tttttaagaa	aaataaatta	atatatgtat	attcttggaa	ctaattgtatg	tggaataaca	9420
gtaacttaat	tattaaacat	tctaaatgca	aatatgcaaa	gaaaaaaaag	aaaagaacac	9480
aactgaaatc	aaagccagat	tcataataat	tggctacatg	gttgtagaat	gtagggtaac	9540
acaacatcca	gaattgaaca	ctcaaattgg	atgatagatg	gataatcttt	agatacaaga	9600
gaattgggtc	tcttccatta	ttaacgaaaa	taaagaaaaa	aagttttagca	taaaagtttg	9660
aaactcaaca	taacattttg	aacttgactc	cttcatagga	gtgacatgaa	ctgacgaatc	9720
acaaccgatt	acttggttga	gtcatcttcc	gctttctcca	ccttcgaaat	gaatgtgacc	9780
ggtttcttcg	ggtgctcatt	tacggtcaag	tgtaaaacat	ctgggtctcga	gtaatgtcca	9840
accgaatcga	agtacaactt	agctcttgct	acatcaccaa	gatcttgatg	ggggatcggg	9900
taccgagctc	gaattcactg	gccgtcggtt	tacaacgact	cagcacgcgt	tggtttcgac	9960
aaaatttaga	acgaacttaa	ttatgatctc	aaatacattg	atacatatct	catctagatc	10020
taggttatca	ttatgtaaga	aagttttgac	gaatatggca	cgacaaaatg	gctagactcg	10080
atgtaattgg	tatctcaact	caacattata	cttataccaa	acattagtta	gacaaaattt	10140
aaacaactat	tttttatgta	tgcaagagtc	agcatatgta	taattgattc	agaatcggtt	10200
tgacgagttc	ggatgtagta	gtagccatta	tttaatgtac	atactaactc	tgaatagtga	10260
atatgatgaa	acattgtatc	ttattgtata	aatatccata	aacacatcat	gaaagacact	10320
ttctttcacg	gtctgaatta	attatgatac	aattctaata	gaaaacgaat	taaattacgt	10380
tgaattgtat	gaaatcta	tgaacaagcc	aaccacgacg	acgactaacg	ttgcctggat	10440

tgactcgggt	taagttaacc	actaaaaaaaa	cggagctgtc	atgtaacacg	cggatcgagc	10500
aggtcacagt	catgaagcca	tcaaagcaaa	agaactaatc	caagggtga	gatgattaat	10560
tagtttaaaa	attagttaac	acgagggaaa	aggctgtctg	acagccaggt	cacgttatct	10620
ttacctgtgg	tcgaaatgat	tcgtgtctgt	cgattttaat	tatttttttg	aaaggccgaa	10680
aataaagttg	taagagataa	accgcctat	ataaattcat	atattttcct	ctccgctttg	10740
aattgtctcg	ttgtcctcct	cactttcatc	agccgttttg	aatctccggc	gacttgacag	10800
agaagaacaa	ggaagaagac	taagagagaa	agtaagagat	aatccaggag	attcattctc	10860
cgttttgaat	cttcctcaat	ctcatcttct	tccgctcttt	ctttccaagg	taataggaac	10920
tttctggatc	tactttatct	gctggatctc	gatcttggtt	tctcaatttc	cttgagatct	10980
ggaattcgtt	taatttggat	ctgtgaacct	ccactaaatc	ttttggtttt	actagaatcg	11040
atctaagttg	accgatcagt	tagctcgatt	atagctacca	gaatttggct	tgaccttgat	11100
ggagagatcc	atgttcatgt	tacctgggaa	atgatttgta	tatgtgaatt	gaaatctgaa	11160
ctgttgaaat	tagattgaat	ctgaacactg	tcaatgttag	attgaatctg	aacactgttt	11220
aagggttagat	gaagtttgtg	tatagattct	tcgaaacttt	aggatttgta	gtgtcgtacg	11280
ttgaacagaa	agctatttct	gattcaatca	gggtttatct	gactgtattg	aactcttttt	11340
gtgtgtttgc	agctcataaa	aaaaacgcga	acctgcaggc	atggcggcgg	caacaacaac	11400
aacaacaaca	tcttcttcga	tctccttctc	caccaaacca	tctccttctc	cctccaaatc	11460
accattacca	atctccagat	tctccctccc	attctcccta	aaccccaaca	aatcatcctc	11520
ctcctcccg	cgccgcggta	tcaaattccag	ctctccctcc	tccatctccg	ccgtgctcaa	11580
cacaaccacc	aatgtcacia	ccactccctc	tccaaccaa	cctaccaaac	ccgaaacatt	11640
catctcccg	ttcgctccag	atcaaccccc	caaaggcgct	gatatcctcg	tcgaagcttt	11700
agaacgtcaa	ggcgtagaaa	ccgtattcgc	ttaccctgga	ggtgcatcaa	tggagattca	11760
ccaagcctta	accgctctct	cctcaatccg	taacgtcctt	cctcgtcacg	aacaaggagg	11820
tgtattcgca	gcagaaggat	acgctcgatc	ctcaggtaaa	ccaggatatct	gtatagccac	11880
ttcagggtccc	ggagctacaa	atctcgttag	cggattagcc	gatgcgttgt	tagatagtgt	11940
tcctcttgta	gcaatcacag	gacaagtccc	tcgtcgtatg	attggtacag	atgcgtttca	12000
agagactccg	attgttgagg	taacgcgttc	gattacgaag	cataactatc	ttgtgatgga	12060
tggtgaagat	atccctagga	ttattgagga	agctttcttt	ttagctactt	ctggtagacc	12120
tggacctgtt	ttggttgatg	ttcctaaaga	tattcaacaa	cgcttgcgca	ttcctaattg	12180
ggaacagggt	atgagattag	ctggttatat	gtctaggatg	cctaaacctc	cggaagattc	12240
tcatttgagg	cagattgtta	gggttgatttc	tgagtctaag	aagcctgtgt	tgtatgttgg	12300
tggttggttgt	ttgaattcta	gcgatgaatt	gggtagggtt	gttgagctta	cggggatccc	12360
tggtgcgagt	acgttgatgg	ggctgggatc	ttatccttgt	gatgatgagt	tgctcgttaca	12420
tatgcttgga	atgcatggga	ctgtgtatgc	aaattacgct	gtggagcata	gtgatttgtt	12480
gttggcgctt	ggggtgaagg	ttgatgatcg	tgtcacgggt	aagcttgagg	cttttgctag	12540
tagggctaag	attgttcata	ttgataattga	ctcggctgag	attgggaaga	ataagactcc	12600
tcatgtgtct	gtgtgtggtg	atgttaagct	ggctttgcaa	gggatgaata	aggttcttga	12660
gaaccgagcg	gaggagctta	agcttgattt	tggagtttgg	aggaatgagt	tgaacgtaca	12720
gaaacagaag	tttccgttga	gctttaagac	gtttggggaa	gctattcctc	cacagtatgc	12780
gattaaggct	cttgatgagt	tgactgatgg	aaaagccata	ataagtactg	gtgtcgggca	12840
acatcaaatg	tgggcggcgc	agttctacaa	ttacaagaaa	ccaaggcagt	ggctatcatc	12900
aggaggcctt	ggagctatgg	gatttgagct	tcctgtgcgc	attggagcgt	ctgttgctaa	12960
ccctgatcgc	atagttgtgg	atattgacgg	agatggaagc	tttataatga	atgtgcaaga	13020
gctagccact	attcgtgtag	agaatcttcc	agtgaaggta	cttttattaa	acaaccagca	13080
tcttgggcatg	gttatgcaat	gggaagatcg	gttctacaaa	gctaaccgag	ctcacacatt	13140
tctcggggat	ccggctcagg	aggacgagat	attcccgaac	atgttgctgt	ttgcagcagc	13200
ttgcgggatt	ccagcggcga	gggtgacaaa	gaaagcagat	ctccgagaag	ctattcagac	13260
aatgctggat	acaccaggac	cttacctgtt	ggatgtgatt	tgtccgcacc	aagaacatgt	13320
gttgccgatg	atcccgaatg	gtggcacttt	caacgatgtc	ataacggaag	gagatggccg	13380
gattaaatac	tgagagatga	aaccggcctg	gccggcccgg	agtggggagg	cacgatggcc	13440
gctttggtcg	atcgacggga	tcgatcctgc	tttaatgaga	tatgcgagac	gcctatgatc	13500
gcatgatatt	tgctttcaat	tctgttgtgc	acgttgtaaa	aaacctgagc	atgtgtagct	13560
cagatcctta	ccgccggttt	cggttcatct	taatgaatat	atcaccctgt	actatcgtat	13620
ttttatgaat	aatattctcc	gttcaattta	ctgattgtac	cctactactt	atatgtacaa	13680
tattaaaatg	aaaacaatat	attgtgtcga	ataggtttat	agcgacatct	atgatagagc	13740
gccacaataa	caaacaattg	cgttttatta	ttacaaatcc	aatttttaaaa	aaagcggcag	13800
aaccggtcaa	acctaaaaga	ctgattacat	aaatcttatt	caaattttcaa	aaggccccag	13860

```

gggctagtat ctacgacaca ccgagcggcg aactaataac gttcactgaa gggaactccg 13920
gttccccgcc ggcgcgcatg ggtgagattc cttgaagttg agtattggcc gtccgctcta 13980
ccgaaagtta cgggcacccat tcaaccgggt ccagcacggc ggccgggtaa ccgacttgct 14040
gccccgagaa ttatgcagca tttttttggt gtatgtgggc cccaaatgaa gtgcaggtca 14100
aaccttgaca gtgacgacaa atcgttgggc ggggtccaggc cgaattttgc gacaacatgt 14160
cgaggctcag caggatgggc ccag 14184

```

&lt;210&gt; 59

&lt;211&gt; 1011

&lt;212&gt; DNA

&lt;213&gt; Zea mays

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(981)

&lt;223&gt; coding for 5- methlythioribose kinase

&lt;400&gt; 59

```

gca cga gca ctc ctc tcc tct cct ctc gcc ggc gca tcg ccc gac tgt 48
Ala Arg Ala Leu Leu Ser Ser Pro Leu Ala Gly Ala Ser Pro Asp Cys
1 5 10 15

cag tca gcc tca gcc atg gcc gcg gag gag gag cag ggc ttc cgc ccg 96
Gln Ser Ala Ser Ala Met Ala Ala Glu Glu Glu Gln Gly Phe Arg Pro
20 25 30

ctg gac gag tcg tcc ctg ctc gcc tac atc aag gcc acg ccg gcg ctc 144
Leu Asp Glu Ser Ser Leu Leu Ala Tyr Ile Lys Ala Thr Pro Ala Leu
35 40 45

gcc tcc cgc ctc ggc ggc ggt ggc agt cta gac tcc atc gag atc aag 192
Ala Ser Arg Leu Gly Gly Gly Gly Ser Leu Asp Ser Ile Glu Ile Lys
50 55 60

gag gtc ggc gac ggc aac ctc aac ttc gtc tac atc gtg cag tcc gag 240
Glu Val Gly Asp Gly Asn Leu Asn Phe Val Tyr Ile Val Gln Ser Glu
65 70 75 80

gcc ggc gcc atc gtc gtc aag cag gcg ctc ccg tac gtg cgc tgc gtg 288
Ala Gly Ala Ile Val Val Lys Gln Ala Leu Pro Tyr Val Arg Cys Val
85 90 95

ggg gat tcg tgg ccc atg acg cgg gag cgc gcc tac ttc gag gcc tcc 336
Gly Asp Ser Trp Pro Met Thr Arg Glu Arg Ala Tyr Phe Glu Ala Ser
100 105 110

acg ctg cgg gag cac ggc cgc ctg tgc ccg gag cac acc ccc gag gtg 384
Thr Leu Arg Glu His Gly Arg Leu Cys Pro Glu His Thr Pro Glu Val
115 120 125

tac cac ttc gac cgg acc ttg tcg ctg atg ggg atg cgc tac atc gag 432
Tyr His Phe Asp Arg Thr Leu Ser Leu Met Gly Met Arg Tyr Ile Glu
130 135 140

ccc ccg cac atc atc ctc cgc aag ggc ctc gtc gcc ggt gtc gag tac 480
Pro Pro His Ile Ile Leu Arg Lys Gly Leu Val Ala Gly Val Glu Tyr
145 150 155 160

ccg ctg ctc gcc gac cac atg tcc gat tac atg gcc aag acg ctc ttc 528
Pro Leu Leu Ala Asp His Met Ser Asp Tyr Met Ala Lys Thr Leu Phe
165 170 175

ttc acc tcc ctc ctc tat aac aat acc acg gat cat aag aac gga gtt 576
Phe Thr Ser Leu Leu Tyr Asn Asn Thr Thr Asp His Lys Asn Gly Val
180 185 190

```

gct aag tac tct gcg aac gtg gag atg tgt agg ctc acg gag caa gtt 624  
 Ala Lys Tyr Ser Ala Asn Val Glu Met Cys Arg Leu Thr Glu Gln Val  
 195 200 205  
 gtg ttc tcg gac cca tac cgt gtt tcc aaa ttt aat cgg tgg acc tcg 672  
 Val Phe Ser Asp Pro Tyr Arg Val Ser Lys Phe Asn Arg Trp Thr Ser  
 210 215 220  
 cct tat ctc gac aaa gat gct gag gca gtt cgc gag gat gat gag ctc 720  
 Pro Tyr Leu Asp Lys Asp Ala Glu Ala Val Arg Glu Asp Asp Glu Leu  
 225 230 235 240  
 aag ttg gaa gta gct ggg ctg aaa tcg atg ttt atc gag aga gct caa 768  
 Lys Leu Glu Val Ala Gly Leu Lys Ser Met Phe Ile Glu Arg Ala Gln  
 245 250 255  
 gct ctg att cat gga gat ctc cac act ggt tct atc atg gtg acc gaa 816  
 Ala Leu Ile His Gly Asp Leu His Thr Gly Ser Ile Met Val Thr Glu  
 260 265 270  
 gtt caa ctc aag tca ttg atc cag aat ttg ggt tct atg ggg cca atg 864  
 Val Gln Leu Lys Ser Leu Ile Gln Asn Leu Gly Ser Met Gly Pro Met  
 275 280 285  
 ggg ttt gat att ggg agc ctt cct tgg aaa cct gat ttt ggg cat act 912  
 Gly Phe Asp Ile Gly Ser Leu Pro Trp Lys Pro Asp Phe Gly His Thr  
 290 295 300  
 atg cac aga atg ggc atg ctg atc aag cga atg atc gta agg ctt aca 960  
 Met His Arg Met Gly Met Leu Ile Lys Arg Met Ile Val Arg Leu Thr  
 305 310 315 320  
 aga atg gat ctt gaa gac aat tgaagagtcg tggaatttgt tccacaaaaa 1011  
 Arg Met Asp Leu Glu Asp Asn  
 325

&lt;210&gt; 60

&lt;211&gt; 327

&lt;212&gt; PRT

&lt;213&gt; Zea mays

&lt;400&gt; 60

Ala Arg Ala Leu Leu Ser Ser Pro Leu Ala Gly Ala Ser Pro Asp Cys  
 1 5 10 15  
 Gln Ser Ala Ser Ala Met Ala Ala Glu Glu Glu Gln Gly Phe Arg Pro  
 20 25 30  
 Leu Asp Glu Ser Ser Leu Leu Ala Tyr Ile Lys Ala Thr Pro Ala Leu  
 35 40 45  
 Ala Ser Arg Leu Gly Gly Gly Gly Ser Leu Asp Ser Ile Glu Ile Lys  
 50 55 60  
 Glu Val Gly Asp Gly Asn Leu Asn Phe Val Tyr Ile Val Gln Ser Glu  
 65 70 75 80  
 Ala Gly Ala Ile Val Val Lys Gln Ala Leu Pro Tyr Val Arg Cys Val  
 85 90 95  
 Gly Asp Ser Trp Pro Met Thr Arg Glu Arg Ala Tyr Phe Glu Ala Ser  
 100 105 110  
 Thr Leu Arg Glu His Gly Arg Leu Cys Pro Glu His Thr Pro Glu Val  
 115 120 125  
 Tyr His Phe Asp Arg Thr Leu Ser Leu Met Gly Met Arg Tyr Ile Glu  
 130 135 140

Pro Pro His Ile Ile Leu Arg Lys Gly Leu Val Ala Gly Val Glu Tyr  
 145 150 155 160  
 Pro Leu Leu Ala Asp His Met Ser Asp Tyr Met Ala Lys Thr Leu Phe  
 165 170 175  
 Phe Thr Ser Leu Leu Tyr Asn Asn Thr Thr Asp His Lys Asn Gly Val  
 180 185 190  
 Ala Lys Tyr Ser Ala Asn Val Glu Met Cys Arg Leu Thr Glu Gln Val  
 195 200 205  
 Val Phe Ser Asp Pro Tyr Arg Val Ser Lys Phe Asn Arg Trp Thr Ser  
 210 215 220  
 Pro Tyr Leu Asp Lys Asp Ala Glu Ala Val Arg Glu Asp Asp Glu Leu  
 225 230 235 240  
 Lys Leu Glu Val Ala Gly Leu Lys Ser Met Phe Ile Glu Arg Ala Gln  
 245 250 255  
 Ala Leu Ile His Gly Asp Leu His Thr Gly Ser Ile Met Val Thr Glu  
 260 265 270  
 Val Gln Leu Lys Ser Leu Ile Gln Asn Leu Gly Ser Met Gly Pro Met  
 275 280 285  
 Gly Phe Asp Ile Gly Ser Leu Pro Trp Lys Pro Asp Phe Gly His Thr  
 290 295 300  
 Met His Arg Met Gly Met Leu Ile Lys Arg Met Ile Val Arg Leu Thr  
 305 310 315 320  
 Arg Met Asp Leu Glu Asp Asn  
 325

&lt;210&gt; 61

&lt;211&gt; 471

&lt;212&gt; DNA

&lt;213&gt; Brassica napus

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (2)..(469)

&lt;223&gt; coding for 5- methlythioribose kinase

&lt;400&gt; 61

a ttt ccg ggt cga cga ttt cgt ggc aat ctc aac ttc gtt ttc atc gtc 49

Phe Pro Gly Arg Arg Phe Arg Gly Asn Leu Asn Phe Val Phe Ile Val

1 5 10 15

atc gga tcc act ggc tca ctc gtc atc aaa cag gcg ctt ccg tat ata 97

Ile Gly Ser Thr Gly Ser Leu Val Ile Lys Gln Ala Leu Pro Tyr Ile

20 25 30

cgt tgt att ggg gag tct tgg cca atg acg aaa gaa aga gct tac ttt 145

Arg Cys Ile Gly Glu Ser Trp Pro Met Thr Lys Glu Arg Ala Tyr Phe

35 40 45

gaa gct aca act ctg aga aag cac gga gct ttg tct cct gat cat gtt 193

Glu Ala Thr Thr Leu Arg Lys His Gly Ala Leu Ser Pro Asp His Val

50 55 60

cct gaa gtc tac cat ttt gac agg acc atg gct ttg att gga atg agg 241

Pro Glu Val Tyr His Phe Asp Arg Thr Met Ala Leu Ile Gly Met Arg

65 70 75 80



tat ctg gag cct cct cac atc atc ctc cgc aaa gga ctc gtt gct gga 289  
 Tyr Leu Glu Pro Pro His Ile Ile Leu Arg Lys Gly Leu Val Ala Gly  
                     85                    90                    95  
 atc cag tac cct ttc ctt gca gaa cac atg gct gat tac atg gcc aaa 337  
 Ile Gln Tyr Pro Phe Leu Ala Glu His Met Ala Asp Tyr Met Ala Lys  
                     100                    105                    110  
 acc ctc ttc ttc act tcg ctc ctc tat cat gat acc aca gag cac aaa 385  
 Thr Leu Phe Phe Thr Ser Leu Leu Tyr His Asp Thr Thr Glu His Lys  
                     115                    120                    125  
 aga gca gta acc gag ttt tgt ggt aat gtg gag tta tgc cgg tta acg 433  
 Arg Ala Val Thr Glu Phe Cys Gly Asn Val Glu Leu Cys Arg Leu Thr  
                     130                    135                    140  
 gag caa gta gtg ttc tct gac ccg tat aga gtt tct ag 471  
 Glu Gln Val Val Phe Ser Asp Pro Tyr Arg Val Ser  
 145                    150                    155

&lt;210&gt; 62

&lt;211&gt; 156

&lt;212&gt; PRT

&lt;213&gt; Brassica napus

&lt;400&gt; 62

Phe Pro Gly Arg Arg Phe Arg Gly Asn Leu Asn Phe Val Phe Ile Val  
 1                    5                    10                    15  
 Ile Gly Ser Thr Gly Ser Leu Val Ile Lys Gln Ala Leu Pro Tyr Ile  
                     20                    25                    30  
 Arg Cys Ile Gly Glu Ser Trp Pro Met Thr Lys Glu Arg Ala Tyr Phe  
                     35                    40                    45  
 Glu Ala Thr Thr Leu Arg Lys His Gly Ala Leu Ser Pro Asp His Val  
                     50                    55                    60  
 Pro Glu Val Tyr His Phe Asp Arg Thr Met Ala Leu Ile Gly Met Arg  
                     65                    70                    75                    80  
 Tyr Leu Glu Pro Pro His Ile Ile Leu Arg Lys Gly Leu Val Ala Gly  
                     85                    90                    95  
 Ile Gln Tyr Pro Phe Leu Ala Glu His Met Ala Asp Tyr Met Ala Lys  
                     100                    105                    110  
 Thr Leu Phe Phe Thr Ser Leu Leu Tyr His Asp Thr Thr Glu His Lys  
                     115                    120                    125  
 Arg Ala Val Thr Glu Phe Cys Gly Asn Val Glu Leu Cys Arg Leu Thr  
                     130                    135                    140  
 Glu Gln Val Val Phe Ser Asp Pro Tyr Arg Val Ser  
 145                    150                    155

&lt;210&gt; 63

&lt;211&gt; 415

&lt;212&gt; DNA

&lt;213&gt; Brassica napus

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (3)..(413)

&lt;223&gt; coding for 5- methlythioribose kinase

<400> 63  
gg gtc gac gat ttc gtg ctg aga gca aaa gag atg tcg ttc gat gag 47  
Val Asp Asp Phe Val Leu Arg Ala Lys Glu Met Ser Phe Asp Glu  
1 5 10 15  
ttc aag ccg ttg aac gag aaa tct cta gta gag tac ata aag gca acg 95  
Phe Lys Pro Leu Asn Glu Lys Ser Leu Val Glu Tyr Ile Lys Ala Thr  
20 25 30  
cct gcc ctc tcc tcc agg ctc gga gac aag tac gat gat ctg gtc atc 143  
Pro Ala Leu Ser Ser Arg Leu Gly Asp Lys Tyr Asp Asp Leu Val Ile  
35 40 45  
aag gaa gtt gga gat ggc aat ctc aac ttc gtt ttc atc gtt gtc gga 191  
Lys Glu Val Gly Asp Gly Asn Leu Asn Phe Val Phe Ile Val Val Gly  
50 55 60  
tcc act ggc tca ctc gtc atc aaa cag gcg ctt ccg tat ata cgt tgt 239  
Ser Thr Gly Ser Leu Val Ile Lys Gln Ala Leu Pro Tyr Ile Arg Cys  
65 70 75  
att gga gaa tca tgg cca atg acg aaa gaa aga gct tac ttt gaa gca 287  
Ile Gly Glu Ser Trp Pro Met Thr Lys Glu Arg Ala Tyr Phe Glu Ala  
80 85 90 95  
aca act ctg aga aag cac ggt ggt ttg tct ccg gat cat gtt cct gaa 335  
Thr Thr Leu Arg Lys His Gly Gly Leu Ser Pro Asp His Val Pro Glu  
100 105 110  
gtc tac cat ttt gac aga acc atg gct ttg att gga atg aga tac ctc 383  
Val Tyr His Phe Asp Arg Thr Met Ala Leu Ile Gly Met Arg Tyr Leu  
115 120 125  
gag cct cct cac atc atc ctc cgc aaa gga ct 415  
Glu Pro Pro His Ile Ile Leu Arg Lys Gly  
130 135

<210> 64

<211> 137

<212> PRT

<213> Brassica napus

<400> 64

Val Asp Asp Phe Val Leu Arg Ala Lys Glu Met Ser Phe Asp Glu Phe  
1 5 10 15  
Lys Pro Leu Asn Glu Lys Ser Leu Val Glu Tyr Ile Lys Ala Thr Pro  
20 25 30  
Ala Leu Ser Ser Arg Leu Gly Asp Lys Tyr Asp Asp Leu Val Ile Lys  
35 40 45  
Glu Val Gly Asp Gly Asn Leu Asn Phe Val Phe Ile Val Val Gly Ser  
50 55 60  
Thr Gly Ser Leu Val Ile Lys Gln Ala Leu Pro Tyr Ile Arg Cys Ile  
65 70 75 80  
Gly Glu Ser Trp Pro Met Thr Lys Glu Arg Ala Tyr Phe Glu Ala Thr  
85 90 95  
Thr Leu Arg Lys His Gly Gly Leu Ser Pro Asp His Val Pro Glu Val  
100 105 110  
Tyr His Phe Asp Arg Thr Met Ala Leu Ile Gly Met Arg Tyr Leu Glu  
115 120 125  
Pro Pro His Ile Ile Leu Arg Lys Gly  
130 135

<210> 65  
 <211> 424  
 <212> DNA  
 <213> Oryza sativa  
 <220>  
 <221> CDS  
 <222> (3)..(422)  
 <223> coding for 5- methlythioribose kinase

<400> 65  
 cc ctt ctc tac aac tcc acc act gat cac aag aaa gga gtt gct cag 47  
 Leu Leu Tyr Asn Ser Thr Thr Asp His Lys Lys Gly Val Ala Gln  
 1 5 10 15  
 tac tgc gat aat gtg gag atg tgt agg ctc aca gag caa gtc gtg ttc 95  
 Tyr Cys Asp Asn Val Glu Met Cys Arg Leu Thr Glu Gln Val Val Phe  
 20 25 30  
 tca gac cca tac atg ctc gcc aaa tac aat cgt tgc aca tca ccc ttc 143  
 Ser Asp Pro Tyr Met Leu Ala Lys Tyr Asn Arg Cys Thr Ser Pro Phe  
 35 40 45  
 cta gat aat gat gct gca gcg gtt cga gag gat gct gag ctt aaa ttg 191  
 Leu Asp Asn Asp Ala Ala Ala Val Arg Glu Asp Ala Glu Leu Lys Leu  
 50 55 60  
 gag att gct gaa ttg aaa tca atg ttt att gag aga gca cag gct ctt 239  
 Glu Ile Ala Glu Leu Lys Ser Met Phe Ile Glu Arg Ala Gln Ala Leu  
 65 70 75  
 ctt cat gga gat ctc cac act ggt tcc atc atg gtg aca cca gat tct 287  
 Leu His Gly Asp Leu His Thr Gly Ser Ile Met Val Thr Pro Asp Ser  
 80 85 90 95  
 act caa gtg att gat cca gaa ttt gct ttc tat ggc cca atg ggt tac 335  
 Thr Gln Val Ile Asp Pro Glu Phe Ala Phe Tyr Gly Pro Met Gly Tyr  
 100 105 110  
 gac att ggg gcc ttc ctg ggg aac ttg att ttg gca tat ttt tca caa 383  
 Asp Ile Gly Ala Phe Leu Gly Asn Leu Ile Leu Ala Tyr Phe Ser Gln  
 115 120 125  
 gat gga cac gct gat caa gca aat gat cgt aag gct tac aa 424  
 Asp Gly His Ala Asp Gln Ala Asn Asp Arg Lys Ala Tyr  
 130 135 140

<210> 66  
 <211> 140  
 <212> PRT  
 <213> Oryza sativa

<400> 66  
 Leu Leu Tyr Asn Ser Thr Thr Asp His Lys Lys Gly Val Ala Gln Tyr  
 1 5 10 15  
 Cys Asp Asn Val Glu Met Cys Arg Leu Thr Glu Gln Val Val Phe Ser  
 20 25 30  
 Asp Pro Tyr Met Leu Ala Lys Tyr Asn Arg Cys Thr Ser Pro Phe Leu  
 35 40 45  
 Asp Asn Asp Ala Ala Ala Val Arg Glu Asp Ala Glu Leu Lys Leu Glu  
 50 55 60

```

Ile Ala Glu Leu Lys Ser Met Phe Ile Glu Arg Ala Gln Ala Leu Leu
 65              70              75              80
His Gly Asp Leu His Thr Gly Ser Ile Met Val Thr Pro Asp Ser Thr
              85              90              95
Gln Val Ile Asp Pro Glu Phe Ala Phe Tyr Gly Pro Met Gly Tyr Asp
              100             105             110
Ile Gly Ala Phe Leu Gly Asn Leu Ile Leu Ala Tyr Phe Ser Gln Asp
              115             120             125
Gly His Ala Asp Gln Ala Asn Asp Arg Lys Ala Tyr
 130              135              140

```

&lt;210&gt; 67

&lt;211&gt; 404

&lt;212&gt; DNA

&lt;213&gt; Glycine max

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (3)..(404)

&lt;223&gt; coding for 5- methlythioribose kinase

&lt;400&gt; 67

```

ta atc ccc gaa cat gtt cct gaa gtg tat cac ttt gac cgt acc atg      47
  Ile Pro Glu His Val Pro Glu Val Tyr His Phe Asp Arg Thr Met
    1              5              10              15
tct ttg atc ggt atg cgt tac ttg gag ccc cca cat ata atc ctc ata      95
Ser Leu Ile Gly Met Arg Tyr Leu Glu Pro Pro His Ile Ile Leu Ile
              20              25              30
aaa ggg ttg att gct ggg att gag tac cct ttt ttg gct gaa cac atg      143
Lys Gly Leu Ile Ala Gly Ile Glu Tyr Pro Phe Leu Ala Glu His Met
              35              40              45
gct gat ttc atg gcg aag aca ctc ttc ttc acg tct ctg ctt ttc cgt      191
Ala Asp Phe Met Ala Lys Thr Leu Phe Phe Thr Ser Leu Leu Phe Arg
              50              55              60
tcc act gct gac cac aaa cgg gac gtt gcc gaa ttt tgt ggg aat gtg      239
Ser Thr Ala Asp His Lys Arg Asp Val Ala Glu Phe Cys Gly Asn Val
              65              70              75
gag tta tgc agg ctc act gaa cag gtc gtt ttc tct gac cct tat aaa      287
Glu Leu Cys Arg Leu Thr Glu Gln Val Val Phe Ser Asp Pro Tyr Lys
              80              85              90              95
gtt tct caa tat aat cgt tgg act tcc ccc tat ctt gat cgt gat gct      335
Val Ser Gln Tyr Asn Arg Trp Thr Ser Pro Tyr Leu Asp Arg Asp Ala
              100             105             110
gag gct gtt cgg gaa gac aat ctg ctg aag ctt gaa gtt gct gag ctg      383
Glu Ala Val Arg Glu Asp Asn Leu Leu Lys Leu Glu Val Ala Glu Leu
              115             120             125
aaa tcc aag ttc att gag agc      404
Lys Ser Lys Phe Ile Glu Ser
              130

```

&lt;210&gt; 68

&lt;211&gt; 134

&lt;212&gt; PRT

&lt;213&gt; Glycine max

&lt;400&gt; 68

Ile Pro Glu His Val Pro Glu Val Tyr His Phe Asp Arg Thr Met Ser  
 1 5 10 15  
 Leu Ile Gly Met Arg Tyr Leu Glu Pro Pro His Ile Ile Leu Ile Lys  
 20 25 30  
 Gly Leu Ile Ala Gly Ile Glu Tyr Pro Phe Leu Ala Glu His Met Ala  
 35 40 45  
 Asp Phe Met Ala Lys Thr Leu Phe Phe Thr Ser Leu Leu Phe Arg Ser  
 50 55 60  
 Thr Ala Asp His Lys Arg Asp Val Ala Glu Phe Cys Gly Asn Val Glu  
 65 70 75 80  
 Leu Cys Arg Leu Thr Glu Gln Val Val Phe Ser Asp Pro Tyr Lys Val  
 85 90 95  
 Ser Gln Tyr Asn Arg Trp Thr Ser Pro Tyr Leu Asp Arg Asp Ala Glu  
 100 105 110  
 Ala Val Arg Glu Asp Asn Leu Leu Lys Leu Glu Val Ala Glu Leu Lys  
 115 120 125  
 Ser Lys Phe Ile Glu Ser  
 130

&lt;210&gt; 69

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
 oligonucleotide primer

&lt;400&gt; 69

cgtgaataacg gcgtggagtc g

21

&lt;210&gt; 70

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
 oligonucleotide primer

&lt;400&gt; 70

cggcaggata atcaggttgg

20

&lt;210&gt; 71

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Künstliche Sequenz

&lt;220&gt;

<223> Beschreibung der künstlichen Sequenz:  
 oligonucleotide primer

&lt;400&gt; 71

gtcaacgtaa ccaaccctgc

20

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
12. Februar 2004 (12.02.2004)

PCT

(10) Internationale Veröffentlichungsnummer  
**WO 2004/013333 A3**

(51) Internationale Patentklassifikation<sup>7</sup>: C12N 15/82,  
9/12, 15/54, 15/11, A01H 5/00

(21) Internationales Aktenzeichen: PCT/EP2003/007877

(22) Internationales Anmeldedatum:  
18. Juli 2003 (18.07.2003)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:  
102 34 287.3 26. Juli 2002 (26.07.2002) DE

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme  
von US): BASF PLANT SCIENCE GMBH [DE/DE]; .,  
67056 Ludwigshafen (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): KOCK, Michael  
[DE/DE]; Am Leutbusch 12, 67105 Schifferstadt  
(DE). FRANK, Markus [DE/DE]; Rheindammstr.  
30, 68163 Mannheim (DE). BADUR, Ralf [DE/DE];  
Theodor-Storm-Str. 7B, 67117 Limburgerhof (DE).

(74) Anwalt: GOLDSCHIED, Bettina; c/o BASF Aktiengesellschaft,  
67056 Ludwigshafen (DE).

(81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,  
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,  
CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE,  
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,  
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,  
MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT,  
RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR,  
TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Bestimmungsstaaten (regional): ARIPO Patent (GH,  
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ,  
TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE,  
DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL,  
PT, RO, SE, SI, SK, TR), OAPI Patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— mit internationalem Recherchenbericht

(88) Veröffentlichungsdatum des internationalen  
Recherchenberichts: 15. Juli 2004

Zur Erklärung der Zweibuchstaben-Codes und der anderen Ab-  
kürzungen wird auf die Erklärungen ("Guidance Notes on Co-  
des and Abbreviations") am Anfang jeder regulären Ausgabe der  
PCT-Gazette verwiesen.

(54) Title: INVERSION OF THE NEGATIVE-SELECTIVE EFFECT OF NEGATIVE MARKER PROTEINS USING SELEC-  
TION METHODS

(54) Bezeichnung: REVERTIERUNG DER NEGATIV-SELEKTIVEN WIRKUNG VON NEGATIVEN MARKERPROTEINEN  
ALS SELEKTIONSVERFAHREN

(57) Abstract: The invention relates to methods for producing transformed plant cells or organisms by transforming a population  
of plant cells comprising at least one marker protein having a directly or indirectly toxic effect therefor, by means of at least one  
nucleic acid sequence to be inserted, said sequence being combined with at least one compound preferably a DNA construct which  
is able to reduce the expression, quantity, activity and/or function of the marker protein. The transformed plant cells have a growth  
advantage in relation to the non-transformed cells as a result of the action of said compound.

(57) Zusammenfassung: Die vorliegende Erfindung betrifft Verfahren zur Herstellung transformierter pflanzlicher Zellen oder Or-  
ganismen durch Transformation einer Population pflanzlicher Zellen, die mindestens ein Markerprotein mit einem für diese direkt  
oder indirekt toxischen Effekt umfasst, mit mindestens einer zu insertierenden Nukleinsäuresequenz in Kombination mit mindes-  
tens einer Verbindung - bevorzugt einem DNA-Konstrukt - befähigt zur Verminderung der Expression, Menge, Aktivität und/oder  
Funktion des Markerproteins, wobei die transformierten pflanzlichen Zellen infolge der Wirkung besagter Verbindung gegenüber  
nicht-transformierten Zellen einen Wachstumsvorteil haben.



WO 2004/013333 A3

## INTERNATIONAL SEARCH REPORT

International Application No

PC 03/07877

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/82 C12N9/12 C12N15/54 C12N15/11 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>GLEAVE A P ET AL: "SELECTABLE MARKER-FREE TRANSGENIC PLANTS WITHOUT SEXUAL CROSSING: TRANSIENT EXPRESSION OF CRE RECOMBINASE AND USE OF A CONDITIONAL LETHAL DOMINANT GENE"</p> <p>PLANT MOLECULAR BIOLOGY, NIJHOFF PUBLISHERS, DORDRECHT, NL, vol. 40, May 1999 (1999-05), pages 223-235, XP000995562</p> <p>ISSN: 0167-4412</p> <p>the whole document</p> <p>-----</p> <p>-/--</p>	1-10,12, 17,18



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

31 October 2003

Date of mailing of the international search report

16.04.04

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Bilang, J

## INTERNATIONAL SEARCH REPORT

International Application No

PC

03/07877

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CORNEILLE S ET AL: "Efficient elimination of selectable marker genes from the plastid genome by the CRE-lox site-specific recombination system" PLANT JOURNAL, BLACKWELL SCIENTIFIC PUBLICATIONS, OXFORD, GB, vol. 27, no. 2, July 2001 (2001-07), pages 171-178, XP002236641 ISSN: 0960-7412 the whole document	1-10,12, 17,18
X	----- RISSEUW EDDY ET AL: "Gene targeting and instability of Agrobacterium T-DNA loci in the plant genome" PLANT JOURNAL, vol. 11, no. 4, 1997, pages 717-728, XP002259897 ISSN: 0960-7412 the whole document	1-10,12, 17,18
A	----- EP 0 716 147 A (JUJO PAPER CO LTD) 12 June 1996 (1996-06-12) page 5, line 26 - page 7, line 25	
A	----- CHUANG CHIOU-FEN ET AL: "Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 97, no. 9, 25 April 2000 (2000-04-25), pages 4985-4990, XP002184942 ISSN: 0027-8424 the whole document	
A	----- SALOMON SIEGFRIED ET AL: "Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells" EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 17, no. 20, 15 October 1998 (1998-10-15), pages 6086-6095, XP002259898 ISSN: 0261-4189 the whole document	
	----- -/--	



## INTERNATIONAL SEARCH REPORT

International Application No

PCT, ... 03/07877

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>ENDO S ET AL: "A new GST-MAT vector containing both ipt and iaaM/H genes can produce marker-free transgenic tobacco plants with high frequency" PLANT CELL REPORTS, vol. 20, no. 10, March 2002 (2002-03), pages 923-928, XP002258506 ISSN: 0721-7714 the whole document -----</p>	

# INTERNATIONAL SEARCH REPORT

International application No.

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

### SEE SUPPLEMENTAL BOX

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-19

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

The International Searching Authority has determined that this international application contains multiple (groups of) inventions, namely

1. Claims 1-19

method of preparing transformed plant cells and organisms, including the transformation of a population of cells that contain a negative marker protein, with a nucleic acid sequence that reduces the effect of the negative marker protein.

---

2. Claims 20, 21

amino acid and nucleic acid sequences coding for a plant 5-methyl-thioribose kinase.

---

3. Claims 22-25 (in full), 28-31 (in part)

double-stranded RNA molecule comprising sense and antisense strands of a marker protein as well as plants containing said molecule.

---

4. Claims 26, 27 (in full), 28-31 (in part)

expression cassette containing a marker protein in the antisense orientation as well as plants that contain said cassette.

---

# INTERNATIONAL SEARCH REPORT

International Application No

PC 03/07877

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0716147	A	12-06-1996	JP 3256952 B2 18-02-2002
			JP 9154580 A 17-06-1997
			AU 703485 B2 25-03-1999
			AU 3855795 A 06-06-1996
			BG 62892 B1 31-10-2000
			BG 101524 A 30-01-1998
			BR 9509715 A 28-10-1997
			CA 2162449 A1 10-05-1996
			CN 1137565 A ,B 11-12-1996
			CZ 9701388 A3 18-02-1998
			EP 0716147 A2 12-06-1996
			FI 971961 A 07-07-1997
			HU 77074 A2 02-03-1998
			WO 9615252 A2 23-05-1996
			JP 2002165531 A 11-06-2002
			NO 972108 A 07-07-1997
			NZ 295256 A 29-04-1999
			PL 320201 A1 15-09-1997
			RU 2149187 C1 20-05-2000
			SK 56997 A3 06-05-1998
			TW 446539 B 21-07-2001
			US 5965791 A 12-10-1999
			ZA 9509485 A 27-02-1997

## INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 03/07877

## A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

IPK 7 C12N15/82 C12N9/12 C12N15/54 C12N15/11 A01H5/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

## B. RECHERCHIERTE GEBIETE

Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 7 C12N

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data, BIOSIS

## C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	GLEAVE A P ET AL: "SELECTABLE MARKER-FREE TRANSGENIC PLANTS WITHOUT SEXUAL CROSSING: TRANSIENT EXPRESSION OF CRE RECOMBINASE AND USE OF A CONDITIONAL LETHAL DOMINANT GENE" PLANT MOLECULAR BIOLOGY, NIJHOFF PUBLISHERS, DORDRECHT, NL, Bd. 40, Mai 1999 (1999-05), Seiten 223-235, XP000995562 ISSN: 0167-4412 das ganze Dokument ----- -/--	1-10, 12, 17, 18

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen☒ Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

- \*A\* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- \*E\* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- \*L\* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- \*O\* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- \*P\* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

\*T\* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

\*X\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderscher Tätigkeit beruhend betrachtet werden

\*Y\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderscher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

\*Z\* Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

31. Oktober 2003

Absenddatum des internationalen Recherchenberichts

16.04.04

Name und Postanschrift der internationalen Recherchenbehörde  
Europäisches Patentamt, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Billang, J

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen  
PCT/EP 03/07877

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	CORNEILLE S ET AL: "Efficient elimination of selectable marker genes from the plastid genome by the CRE-lox site-specific recombination system" PLANT JOURNAL, BLACKWELL SCIENTIFIC PUBLICATIONS, OXFORD, GB, Bd. 27, Nr. 2, Juli 2001 (2001-07), Seiten 171-178, XP002236641 ISSN: 0960-7412 das ganze Dokument	1-10,12, 17,18
X	RISSEEUW EDDY ET AL: "Gene targeting and instability of Agrobacterium T-DNA loci in the plant genome" PLANT JOURNAL, Bd. 11, Nr. 4, 1997, Seiten 717-728, XP002259897 ISSN: 0960-7412 das ganze Dokument	1-10,12, 17,18
A	EP 0 716 147 A (JUJO PAPER CO LTD) 12. Juni 1996 (1996-06-12) Seite 5, Zeile 26 - Seite 7, Zeile 25	
A	CHUANG CHIOU-FEN ET AL: "Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, Bd. 97, Nr. 9, 25. April 2000 (2000-04-25), Seiten 4985-4990, XP002184942 ISSN: 0027-8424 das ganze Dokument	
A	SALOMON SIEGFRIED ET AL: "Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells" EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, Bd. 17, Nr. 20, 15. Oktober 1998 (1998-10-15), Seiten 6086-6095, XP002259898 ISSN: 0261-4189 das ganze Dokument	
	----- -/-	

# INTERNATIONALER RECHERCHENBERICHT

In: inales Aktenzeichen

PCT/EP 03/07877

## C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>ENDO S ET AL: "A new GST-MAT vector containing both ipt and iaaM/H genes can produce marker-free transgenic tobacco plants with high frequency"</p> <p>PLANT CELL REPORTS, Bd. 20, Nr. 10, März 2002 (2002-03), Seiten 923-928, XP002258506 ISSN: 0721-7714 das ganze Dokument</p> <p>-----</p>	

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen  
PCT/EP 03/07877

## Feld I Bemerkungen zu den Ansprüchen, die sich als nicht recherchierbar erwiesen haben (Fortsetzung von Punkt 2 auf Blatt 1)

Gemäß Artikel 17(2)a) wurde aus folgenden Gründen für bestimmte Ansprüche kein Recherchenbericht erstellt:

1. ☐ Ansprüche Nr.  
weil sie sich auf Gegenstände beziehen, zu deren Recherche die Behörde nicht verpflichtet ist, nämlich
2. ☐ Ansprüche Nr.  
weil sie sich auf Teile der internationalen Anmeldung beziehen, die den vorgeschriebenen Anforderungen so wenig entsprechen, daß eine sinnvolle internationale Recherche nicht durchgeführt werden kann, nämlich
3. ☐ Ansprüche Nr.  
weil es sich dabei um abhängige Ansprüche handelt, die nicht entsprechend Satz 2 und 3 der Regel 6.4 a) abgefaßt sind.

## Feld II Bemerkungen bei mangelnder Einheitlichkeit der Erfindung (Fortsetzung von Punkt 3 auf Blatt 1)

Die internationale Recherchenbehörde hat festgestellt, daß diese internationale Anmeldung mehrere Erfindungen enthält:

siehe Zusatzblatt

1. ☐ Da der Anmelder alle erforderlichen zusätzlichen Recherchegebühren rechtzeitig entrichtet hat, erstreckt sich dieser internationale Recherchenbericht auf alle recherchierbaren Ansprüche.
2. ☐ Da für alle recherchierbaren Ansprüche die Recherche ohne einen Arbeitsaufwand durchgeführt werden konnte, der eine zusätzliche Recherchegebühr gerechtfertigt hätte, hat die Behörde nicht zur Zahlung einer solchen Gebühr aufgefordert.
3. ☐ Da der Anmelder nur einige der erforderlichen zusätzlichen Recherchegebühren rechtzeitig entrichtet hat, erstreckt sich dieser internationale Recherchenbericht nur auf die Ansprüche, für die Gebühren entrichtet worden sind, nämlich auf die Ansprüche Nr.
4. ☒ Der Anmelder hat die erforderlichen zusätzlichen Recherchegebühren nicht rechtzeitig entrichtet. Der internationale Recherchenbericht beschränkt sich daher auf die in den Ansprüchen zuerst erwähnte Erfindung; diese ist in folgenden Ansprüchen erfaßt:  
1-19

Bemerkungen hinsichtlich eines Widerspruchs

- ☐ Die zusätzlichen Gebühren wurden vom Anmelder unter Widerspruch gezahlt.
- ☐ Die Zahlung zusätzlicher Recherchegebühren erfolgte ohne Widerspruch.



## WEITERE ANGABEN

PCT/ISA/ 210

Die internationale Recherchenbehörde hat festgestellt, dass diese internationale Anmeldung mehrere (Gruppen von) Erfindungen enthält, nämlich:

## 1. Ansprüche: 1-19

Verfahren zur Herstellung transformierter pflanzlicher Zellen und Organismen, beinhaltend die Transformation einer Population von Zellen, welche ein negatives Markerprotein enthalten, mit einer Nukleinsäuresequenz, welche die Wirkung des negativen Markerproteins reduziert

---

## 2. Ansprüche: 20, 21

Aminosäuren- und Nukleinsäuresequenzen kodierend für eine pflanzliche 5-methyl-thioribosekinase

---

## 3. Ansprüche: 22-25 (ganz), 28-31 (teilweise)

Doppelsträngiges RNA-Molekül umfassend sense- und antisense-Strang eines Markerproteins, sowie Pflanzen, welche besagtes Molekül enthalten

---

## 4. Ansprüche: 26, 27 (ganz), 28-31 (teilweise)

Expressionskassette enthaltend ein Markerprotein in Antisense-Orientierung, sowie Pflanzen, welche besagte Kassette enthalten

---

# INTERNATIONALER RECHERCHENBERICHT

Inter  
ales Aktenzeichen  
PCT/EP 03/07877

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 0716147 A	12-06-1996	JP 3256952 B2	18-02-2002
		JP 9154580 A	17-06-1997
		AU 703485 B2	25-03-1999
		AU 3855795 A	06-06-1996
		BG 62892 B1	31-10-2000
		BG 101524 A	30-01-1998
		BR 9509715 A	28-10-1997
		CA 2162449 A1	10-05-1996
		CN 1137565 A , B	11-12-1996
		CZ 9701388 A3	18-02-1998
		EP 0716147 A2	12-06-1996
		FI 971961 A	07-07-1997
		HU 77074 A2	02-03-1998
		WO 9615252 A2	23-05-1996
		JP 2002165531 A	11-06-2002
		NO 972108 A	07-07-1997
		NZ 295256 A	29-04-1999
		PL 320201 A1	15-09-1997
		RU 2149187 C1	20-05-2000
		SK 56997 A3	06-05-1998
		TW 446539 B	21-07-2001
		US 5965791 A	12-10-1999
		ZA 9509485 A	27-02-1997

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**